PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6: C12N 15/82, 5/10, C07K 14/415, A01H 5/00 — A1

(11) International Publication Number: WO 97/49822

(43) International Publication Date: 31 December 1997 (31.12.97)

(21) International Application Number: PCT/EP97/01218

(22) International Filing Date: 10 March 1997 (10.03.97)

(30) Priority Data:

60/020,272	21 June 1996 (21.06.96)	US
60/024,883	30 August 1996 (30.08.96)	US
60/033,177	13 December 1996 (13.12.96)	US
08/773,559	27 December 1996 (27.12.96)	US
60/035,022	10 January 1997 (10.01.97)	US

(71) Applicant *(for all designated States except US)*: NOVARTIS AG [CH/CH]; Schwarzwaldallee 215, CH-4058 Basel (CH).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: RYALS, John, Andrew [US/US]; 14 Sanderling Court, Durham, NC 27713 (US). DELANEY, Terrence, Patrick [US/US]; 4900 Greenview Drive, Durham, NC 27713 (US). FRIEDRICH, Leslie, Bethards [US/US]; 106 North Woodshed Court, Cary, NC 27513 (US). WEYMANN, Kristianna [US/US]; 7033 S.E. 34th Street, Portland, OR 97202 (US). JOHNSON, Jay, Earl [US/US]; 112 Fox Hunt Court, Youngsville, NC 27596 (US). LAWTON, Kay, Ann [US/US]; 8016 Crooked Chute Court, Raleigh, NC 27612 (US). ELLIS, Daniel, Murray [US/US]; 201 Medcon Court, Cary, NC 27511 (US).

(74) Agent: ROTH, Bernhard, M.; Novartis AG, Patent- und Markenabteilung, Klybeckstrasse 141, CH-4002 Basel (CH).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: GENE CONFERRING DISEASE RESISTANCE IN PLANTS AND USES THEREOF

(57) Abstract

The invention concerns the location and characterization of a gene (designated *NIM1*) which is a key component of the SAR pathway and which in connection with chemical and biological inducers enables induction of SAR gene expression and broad spectrum disease resistance to plants. The invention further concerns plants transformed with the *NIM1* gene as well as methods employing the gene to create the transgenic plants and employing the gene in a screening assay for compounds capable of inducing broad spectrum disease resistance in plants.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

GENE CONFERRING DISEASE RESISTANCE IN PLANTS AND USES THEREOF

The present invention relates to disease resistance in plants and to identifying and breeding disease resistance into plants. More particularly, the present invention relates to the identification, isolation and characterization of a gene involved in broad spectrum disease resistance in plants.

Plants are constantly challenged by a wide variety of pathogenic organisms including viruses, bacteria, fungi, and nematodes. Crop plants are particularly vulnerable because they are usually grown as genetically-uniform monocultures; when disease strikes, losses can be severe.

However, most plants have their own innate mechanisms of defense against pathogenic organisms. Natural variation for resistance to plant pathogens has been identified by plant breeders and pathologists and bred into many crop plants. These natural disease resistance genes often provide high levels of resistance to or immunity against pathogens.

In many plant species an initial inoculation by a necrotizing pathogen can immunize the plant to subsequent infection. This acquired disease resistance was first documented in 1901 and is thought to play an important role in the preservation of plants in nature. Particularly well characterized examples of plant immunity are the phenomenon of systemic acquired resistance (SAR) and induced resistance in plants such as tobacco, *Arabidopsis* and cucumber. In these systems, inoculation by a necrotizing pathogen results in systemic protection against subsequent infections by that pathogen as well as a number of other agronomically important bacterial, fungal and viral pathogens.

Systemic acquired resistance can also be triggered by chemical immunization compounds, certain chemicals that induce the immunity response in plants. Such compounds can be of natural origin, such as salicylic acid (SA), or can be synthetic chemicals, such as 2,6-dichloroisonicotinic acid (INA) and benzo(1,2,3)thiadiazole-7-carbothioic acid S-methyl ester (BTH). Treatment with a pathogen or an immunization compound induces the expression of at least nine sets of genes in tobacco, the best characterized species. Different numbers and types of genes can be expressed in other plants. The level of induction for SAR-related genes induced by immunization compounds is as high as 10,000-fold over background. In particular, SAR is characterized by the expression of SAR genes, including the pathogenesis-related (PR) genes.

The SAR genes are induced following infection by a pathogen. Some of these genes have a role in providing systemic acquired resistance to the plant. These plant proteins are induced in large amounts in response to infection by various pathogens, including viruses, bacteria and fungi. PR proteins were first discovered in tobacco plants (*Nicotiana tabacum*) reacting hypersensitively to infection with tobacco mosaic virus (TMV). Subsequently, PR proteins have been found in many plant species (see Redolfi et al. (1983) Neth J Plant Pathol 89: 245-254; Van Loon (1985) Plant Mol. Biol. 4: 111-116; and Uknes et al. (1992) Plant Cell 4: 645-656.) Such proteins are believed to be a common defensive systemic response of plants to infection by pathogens.

Pathogenesis-related proteins include but are not limited to SAR8.2a and SAR8.2b proteins, the acidic and basic forms of tobacco PR-la, PR-lb, and PR-lc; PR-1', PR-2, PR-2', PR-2", PR-N, PR-O, PR-O', PR-4, PR-P, PR-Q, PR-S, and PR-R major proteins; cucumber peroxidases; basic cucumber peroxidase; the chitinase which is a basic counterpart of PR-P or PR-Q; the beta- 1,3-glucanase (glucan endo- 1,3-beta-glucosidase, EC 3.2.1.39) which is a basic counterpart of PR-2, PR-N or PR-O; and the pathogen-inducible chitinase from

cucumber. Such PR proteins are disclosed, for example, in Uknes et al. (1992) The Plant Cell 4: 645-656 and the references cited therein.

SAR or SAR-like genes are expressed in all plant species exhibiting systemic acquired resistance. Expression of such genes can be determined by probing with known SAR DNA sequences. For example, see Lawton et al. (1992) Proceedings of the Second European Federation of Plant Pathology (1983), In: Mechanisms of Defence Responses in Plants, B. Fritig and M. Legrand (eds), Kluwer Academic Publishers, Dordrecht, pp. 410-420; Uknes et al. (1992) The Plant Cell 4: 645-656; and Ward et al. (1991) The Plant Cell 3: 1085-1094. Methods for hybridization and cloning are well known in the art. See, for example, Molecular Cloning, A Laboratory Manual, 2nd Edition, Vol. 1-3, Sambrook et al. (eds.) Cold Spring Harbor Laboratory Press (1989) and the references cited therein.

Alternatively such SAR or SAR-like genes can be found by other methods such as protein screening, +/- screening, etc. See, for example, Liang and Pardee (1992) Science 257: 967-971; and St. John and Davis (1979) Cell 16: 443.

Despite much research and the use of sophisticated and intensive crop-protection measures, including genetic transformation of plants, losses due to disease remain in the billions of dollars annually. Disease resistance genes have previously been cloned but transgenic plants transformed with these genes would typically be resistant only to a subset of strains of a particular pathogen species. Despite efforts to clone genes involved in SAR, a gene controlling broad spectrum disease resistance has not been isolated and characterized.

Several lines of evidence indicate that endogenously produced SA is involved in the signal transduction pathway coupling the perception of pathogen infection with the onset of SAR. Mutants which retain the ability to accumulate SA in response to pathogen yet have lost the ability to induce SAR genes or resistance after application of SA or INA have been

described by Delaney, et al., Proc. Natl. Acad. Sci. 92: 6602-6606 (1995) and in WO94/16077 the whole of which is incorporated herein by reference.

It has now been discovered that these mutants contain a mutant gene, which gene in its wildtype form controls SAR gene expression and SAR itself. The present invention recognizes that the mutant gene confers broad spectrum disease susceptibility to mutant plants and renders them noninducible to pathogens and chemical inducers.

The present invention concerns the identification, isolation and characterization of the wildtype (*NIM1*) gene, a gene which allows activation in a plant of SAR and SAR gene expression in response to biological and chemical inducers.

A mutant gene has been identified in Mutagenized *Arabidopsis* plants. These plants have been found to be defective in their normal response to pathogen infection in that they do not express genes associated with systemic acquired resistance (SAR) nor are they capable of exhibiting SAR. These mutants contain a defective gene which has been labelled *nim1* (for noninducible immunity).

The present invention also concerns the use of the cloned *NIM1* gene and variants thereof to create transgenic plants that have broad spectrum disease resistance and to the transgenic plants produced thereby. The invention further concerns the use of the cloned *NIM1* gene and variants thereof in a screening method for identifying compounds capable of inducing broad spectrum disease resistance in plants.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the effect of chemical inducers on the induction of PR gene expression in wild-type and *nim1* plants.

Figure 2 depicts PR-1 gene expression in pathogen-infected Ws-O and *nim1* plants over the course of 6 days from the initiation of infection.

Figure 3 shows the levels of SA accumulation in Ws-O and *nim1* plants infected with *P. syringae.*

Figure 4 shows the genetic map of the *NIM1* region as determined by AFLP and SSLP analysis.

Figure 5 depicts a physical map of the *NIM1* region as determined by YAC clone analysis.

Figure 6 shows a physical map of an extended P1/BAC contig.

Figure 7 shows a physical map setting forth the positions of P1 and BAC clones with respect to the flanking AFLP markers and YACs.

Figure 8 shows a physical map of a further extended P1/BAC contig containing the *NIM1* gene.

Figure 9 shows an integrated genetic and physical fine map of the *NIM* region.

Figure 10 shows an integrated map of the *NIM1* region.

Figure 11 shows an integrated map of the *NIM1* region including the new AFLP markers.

Figure 12 is a schematic representation of recombinants D169 and C105.

Figure 13 is a global map of the chromosomal region centered on *NIM1* with recombinants indicated, including, BACs, YACs and Cosmids in *NIM1* region.

Figure 14 provides the sequence of the 9.9kb region of clone BAC-04 containing the *NIM1* gene.

Figure 15 shows the nucleic acid sequence of the *NIM1* gene and the amino acid sequence of the *NIM1* gene product, including changes in the various alleles.

Figure 16 shows the expression of *NIM1* induced by INA, BTH, SA and pathogen in wild type and mutant alleles of *nim1*.

Figure 17 shows the expression of PR-1 in *nim1* mutants and wild-type plants.

Figure 18 shows disease resistance in various *nim1* mutants.

Figure 19 is an amino acid sequence comparison of Expressed Sequence Tag regions of the *NIM1* protein and cDNA protein products of 4 rice gene sequences (see SEQ ID NO: 3).

DEFINITIONS

AA:	Amino Acid
AFLP:	Amplified Fragment Length Polymorphism
avrRpt2:	avirulence gene Rpt2, isolated from *Pseudomonas syringae*
BAC:	Bacterial Artificial Chromosome
BTH:	benzo(1,2,3)thiadiazole-7-carbothioic acid S-methyl ester
Col:	*Arabidopsis* ecotype Columbia
ECs:	Enzyme combinations
INA:	2,6-dichloroisonicotinic acid
Ler:	*Arabidopsis* ecotype *Landsberg erecta*
NIM1:	the wildtype gene, conferring disease resistance to the plant
nim:	mutant allele of *NIM1*, conferring disease susceptibility to the plant
nim1:	mutant plant line
ORF:	open reading frame
PCs:	Primer combinations

SA: salicylic acid

SAR: Systemic Acquired Resistance

SSLP: Simple Sequence Length Polymorphism

Ws-O: *Arabidopsis* ecotype Wassilewskija

YAC: Yeast Artificial Chromosome

The *NIM1* gene has been cloned by mapping and walking techniques which indicate that the gene is contained in a ~105 Kb region. (See Figure 13 and Table 16). This region is delineated by the L84.6b marker on the left and the L84.T2 marker on the right. Only three overlapping cosmids made from wild-type DNA from the 105 Kb region complement the *nim1* mutant phenotype (Figure 13 and Table 16). These three cosmids only overlap in a 9.9 Kb region defined by the left end of cosmid clone D7 and the right end of cosmid D5 as pictured in Figure 13. Many other cosmids made to other areas of the 105 Kb region do not complement the *nim1* phenotype (Figure 13 and Table 16). A near full length cDNA clone to the *NIM1* gene indicates the appropriate intron-exon borders and defines the amino acid sequence of the gene product. Only the *NIM1* gene region within the 9.9 Kb complementing region has sequence changes in various *nim1* mutant alleles (Table 18). Three other potential gene regions showed no sequence changes that are assosciated with the *nim1* phenotype. The sequence changes found in the *NIM1* gene region are consistent with altered function or loss of function of the gene product. The severity of the change to the *NIM1* gene region in a particular mutant allele is roughly correlated with the observed physiological severity of that *nim1* allele. Only the *NIM1* gene region had detectable RNA (transcription) and this RNA showed abundant changes consistent with the physiological role of *NIM1* in pathogenesis (Table 18 and Figure 16).

The present invention relates to an isolated gene fragment, the *NIM1* gene, which is a key component of the systemic acquired resistance (SAR) pathway in plants. The *NIM1* gene is associated with the activation of SAR by chemical and biological inducers and, in conjunction with such inducers, is required for SAR and SAR gene expression.

The location of the *NIM* gene is determined by molecular biological analysis of the genome of mutant plants known to carry the mutant *nim1* gene, which gives the host plants extreme sensitivity to a wide variety of pathogens and renders them unable to respond to pathogens and chemical inducers of SAR.

Nim1 mutants are useful as "universal disease susceptible" (UDS) plants by virtue of their being susceptible to many strains and pathotypes of pathogens of the host plant and also to pathogens which do not normally infect the host plant, but which infect other hosts. They can be generated by treating seeds or other biological material with mutagenic agents and then selecting progeny plants for the UDS phenotype by treating progeny plants with known chemical inducers (e.g. INA) of the systemic acquired response and then infecting the plants with a known pathogen. Noninducible mutants develop severe disease symptoms under these circumstances, whereas non-mutants are induced by the chemical compound to systemic acquired resistance. *nim* mutants can be equally selected from mutant populations generated by chemical and irradiation mutagenesis, as well as from populations generated by T-DNA insertion and transposon-induced mutagenesis.

Techniques for generating mutant plant lines are well known in the art. The *nim* plant phenotype is used as a tool to identify an isolated gene fragment which allows expression of broad spectrum disease resistance in plants.

Comprised by the present invention is an isolated DNA molecule comprising a mutant gene of the *NIM1* gene which is a nim1 gene.

Following the use of a *nim1* mutant or plant to isolate the wild-type *NIM1* gene necessary for constitutive expression of SAR genes, the resistance trait, in combination with other characteristics important for production and quality, can be incorporated into plant lines through breeding. Breeding approaches and techniques are known in the art. See, for example, Welsh J. R., Fundamentals of Plant Genetics and Breeding, John Wiley & Sons, NY (1981); Crop Breeding, Wood D. R. (Ed.) American Society of Agronomy Madison, Wisconsin (1983); Mayo O., The Theory of Plant Breeding, Second Edition, Clarendon Press, Oxford (1987); Singh, D.P., Breeding for Resistance to Diseases and Insect Pests, Springer-Verlag, NY (1986); and Wricke and Weber, Quantitative Genetics and Selection Plant Breeding, Walter de Gruyter and Co., Berlin 1986).

A further object of the invention is a chimeric gene comprising a promotor active in plant operably linked to a heterologuous DNA molecule encoding the aminoacid sequence of a *NIM1* gene product and variants thereof according to the invention.

Methodologies for the construction of plant expression cassettes as well as the introduction of foreign DNA into plants is generally described in the art. Generally, for the introduction of foreign DNA into plants, Ti plasmid vectors have been utilized for the delivery of foreign DNA. Also utilized for such delivery have been direct DNA uptake, liposomes, electroporation, micro-injection, and microprojectiles. Such methods had been published in the art. See, for example, Bilang et al. (1991) Gene 100: 247-250; Scheid et al., (1991) Mol. Gen. Genet. 228: 104-112; Guerche et al., (1987) Plant Science 52: 111-116; Neuhause et al., (1987) Theor. Appl. Genet. 75: 30-36; Klein et al., (1987) Nature 327: 70-73; Howell et al., (1980) Science 208:1265; Horsch et al., (1985) Science 227: 1229-1231; DeBlock et al., (1989) Plant Physiology 91: 694-701; Methods for Plant Molecular Biology (Weissbach and Weissbach, eds.) Academic Press, Inc. (1988); and Methods in Plant Molecular Biology (Schuler and Zielinski, eds.) Academic Press, Inc. (1989). See also US. patent application

Serial Nos. 08/438,666 filed May 10, 1995, and WO 93/07278, both of which are incorporated herein by reference in their entirety. It is understood that the method of transformation will depend upon the plant cell to be transformed.

It is further recognized that the components of the expression cassette may be modified to increase expression. For example, truncated sequences, nucleotide substitutions or other modifications may be employed. Plant cells transformed with such modified expression systems, then, would exhibit overexpression or constitutive expression of SAR genes necessary for activation of SAR.

The DNA molecule or gene fragment conferring disease resistance to plants by allowing induction of SAR gene expression can be incorporated in plant or bacterial cells using conventional recombinant DNA technology. Generally, this involves inserting the DNA molecule into an expression system to which the DNA molecule is heterologous (i.e., not normally present). The heterologous DNA molecule is inserted into the expression system or vector in proper orientation and correct reading frame. The vector contains the necessary elements for the transcription and translation of the inserted protein-coding sequences. A large number of vector systems known in the art can be used, such as plasmids, bacteriophage viruses and other modified viruses. Suitable vectors include, but are not limited to, viral vectors such as lambda vector systems lgtl1, lgtl0 and Charon 4; plasmid vectors such as pBI121, pBR322, pACYC177, pACYC184, pAR series, pKK223-3, pUC8, pUC9, pUC18, pUC19, pLG339, pRK290, pKC37, pKC101, pCDNAII; and other similar systems. The DNA sequences can be cloned into the vector using standard cloning procedures in the art, as described by Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Laboratory, Cold Spring Harbor, New York (1982).

A further object of the invention a recombinant vector comprising the chimeric gene according to the invention.

In order to obtain efficient expression of the gene or gene fragment of the present invention, a promoter must be present in the expression vector. RNA polymerase normally binds to the promoter and initiates transcription of a gene. Promoters vary in their strength, i.e., ability to promote transcription. Depending upon the host cell system utilized, any one of a number of suitable promoters can be used. Suitable promoters include ubiquitin, nos promoter, the small subunit ribulose bisphosphate carboxylase gene promoter, the small subunit chlorophyll A/B binding polypeptide promoter, the 35S promoter of cauliflower mosaic virus, and promoters isolated from plant genes. See C.E. Vallejos, et al., "Localization in the Tomato Genome of DNA Restriction Fragments Containing Sequences Homologous to the $_{R}$RNA (45S), the major chlorophyll $_{A/B}$Binding Polypeptide and the Ribulose Bisphosphate Carboxylase Genes," Genetics 112: 93-105 (1986), which discloses the small subunit materials. The nos promoter and the 35S promoter of cauliflower mosaic virus are well known in the art.

Once the disease resistance gene of the present invention has been cloned into an expression system, it is ready to be transformed into a plant cell. Plant tissues suitable for transformation include leaf tissues, root tissues, meristems, and protoplasts.

Bacteria from the genus *Agrobacterium* can be utilized to transform plant cells. Suitable species of such bacterium include *Agrobacterium tumefaciens* and *Agrobacterium rhizogens*. *Agrobacterium tumefaciens* (e.g., strains LBA4404 or EHA105) is particularly useful due to its well-known ability to transform plants.

Another approach to transforming plant cells with a gene involves propelling inert or biologically active particles at plant tissues and cells. This technique is disclosed in U.S. Patent Nos. 4,945,050; 5,036,006; and 5,100,792 all to Sanford et al. Generally, this procedure involves propelling inert or biologically active particles at the cells under conditions effective to penetrate the outer surface of the cell and afford incorporation within

the interior thereof. When inert particles are utilized, the vector can be introduced into the cell by coating the particles with the vector containing the desired gene. Alternatively, the target cell can be surrounded by the vector so that the vector is carried into the cell by the wake of the particle. Biologically active particles (e.g., dried yeast cells, dried bacterium or a bacteriophage, each containing DNA sought to be introduced) can also be propelled into plant cell tissue.

The isolated gene fragment of the present invention can be utilized to confer disease resistance to a wide variety of plant cells, including those of gymnosperms, monocots, and dicots. Although the gene can be inserted into any plant cell falling within these broad classes, it is particularly useful in crop plant cells, such as rice, wheat, barley, rye, corn, potato, carrot, sweet potato, sugar beet, bean, pea, chicory, lettuce, cabbage, cauliflower, broccoli, turnip, radish, spinach, asparagus, onion, garlic, eggplant, pepper, celery, carrot, squash, pumpkin, zucchini, cucumber, apple, pear, quince, melon, plum, cherry, peach, nectarine, apricot, strawberry, grape, raspberry, blackberry, pineapple, avocado, papaya, mango, banana, soybean, tobacco, tomato, sorghum and sugarcane.

The expression system of the present invention can be used to transform virtually any crop plant cell under suitable conditions. Transformed cells can be regenerated into whole plants such that the gene imparts disease resistance to the intact transgenic plants. As set forth above, the expression system can be modified so that the disease resistance gene is continuously or constitutively expressed.

Transformation

The present system can be utilized in any plant which can be transformed and regenerated. Such methods for transformation and regeneration are well known in the art. As well as the above cited references, see also, An, G., Watson, B.D., and Chiang, C.C.

Transformation of tobacco, tomato, potato, and *Arabidopsis thaliana* using a binary Ti vector system. *Plant Physiol.* 81:301-305, 1986; Fry, J., Barnason, A., and Horsch, R.B. Transformation of Brassica napus with Agrobacterium tumefaciens based vectors. *Pl.Cell Rep.* 6:321-325, 1987; Block, M.d. Genotype independent leaf disc transformation of potato (Solanum tuberosum) using Agrobacterium tumefaciens. *Theor.appl.genet.* 76:767-774, 1988; Deblock, M., Brouwer, D.D., and Tenning, P. Transformation of Brassica napus and Brassica oleracea using Agrobacterium tumefaciens and the Expression of the bar and neo genes in the transgenic plants. *Plant Physiol.* 91 :694-701, 1989; Baribault, T.J., Skene, K.G.M., Cain, P.A., and Scott, N.S. Transgenic grapevines: regeneration of shoots expressing beta-glucuronidase. *Pl.Cell Rep.* 41:1045-1049, 1990; Hinchee, M.A.W., Newell, C.A., ConnorWard, D.V., Armstrong, T.A., Deaton, W.R., Sato, S.S., and Rozman, R.J. Transformation and regeneration of non-solanaceous crop plants. *Stadler.Genet.Symp.* 203212.203-212, 1990; Barfield, D.G. and Pua, E.C. Gene transfer in plants of Brassica juncea using Agrobacterium tumefaciens-mediated transformation. *Pl.Cell Rep.* 10:308-314, 1991; Cousins, Y.L., Lyon, B.R., and Llewellyn, D.J. Transformation of an Australian cotton cultivar: prospects for cotton improvement through genetic engineering. *Aust. J.Plant Physiol.* 18:481 -494. 1991; Chee, P.P. and Slightom, J.L. Transformation of Cucumber Tissues by Microprojectile Bombardment Identification of Plants Containing Functional and Nonfunctional Transferred Genes. *GENE* 118:255-260, 1992; Christou, P., Ford, T.L., and Kofron, M. The development of a variety-independent gene-transfer method for rice. *Trends.Biotechnol.* 10:239-246, 1992; D'Halluin, K., Bossut, M., Bonne, E., Mazur, B., Leemans, J., and Botterman, J. Transformation of sugarbeet (Beta vulgaris L.) and evaluation of herbicide resistance in transgenic plants. *Bio/Technol.* 10:309-314. 1992; Dhir, S.K., Dhir, S., Savka, M.A., Belanger, F., Kriz, A.L., Farrand, S.K., and Widholm, J.M. Regeneration of Transgenic Soybean (Glycine Max) Plants from Electroporated Protoplasts.

PLANT PHYSIOL 99:81-88, 1992; Ha, S.B., Wu, F.S., and Thorne, T. K. Transgenic turf-type tall fescue (Festuca arundinacea Schreb.) plants regenerated from protoplasts. *Pl. Cell Rep.* 11:601 -604, 1992; Blechl, A.E. Genetic Transformation The New Tool for Wheat Improvement 78th Annual Meeting Keynote Address. *CEREAL FOOD WORLD* 38:846-847, 1993; Casas, A.M., Kononowicz, A.K., Zehr, U.B., Tomes, D.T., Axtell, J.D., Butler, L.G., Bressan, R.A., and Hasegawa, P.M. Transgenic Sorghum Plants via Microprojectile Bombardment. *PROC NAT ACAD SCI USA 90:11212*-11216, 1993; Christou, P. Philosophy and Practice of Variety Independent Gene Transfer into Recalcitrant Crops. *IN VITRO CELL DEV BIOL-PLANT* 29P: 119-124, 1993; Damiani, F., Nenz, E., Paolocci, F., and Arcioni, S. Introduction of Hygromycin Resistance in Lotus spp Through Agrobacterium Rhizogenes Transformation. *TRANSGENIC RES* 2:330-335, 1993; Davies, D.R., Hamilton, J., and Mullineaux, P. Transformation of Peas. *Pl.Cell Rep.* 12:180-183, 1993; Dong, J.Z. and Mchughen, A. Transgenic Flax Plants from Agrobacterium Mediated Transformation Incidence of Chimeric Regenerants and Inheritance of Transgenic Plants. *PLANT SCI* 91:139-148, 1993; Fitch, M.M.M., Manshardt, R.M., Gonsalves, D., and Slightom, J.L. Transgenic Papaya Plants from Agrobacterium Mediated Transformation of Somatic Embryos. *Pl.Cell Rep.* 12:245-249, 1993; Franklin, C.I. and Trieu, T.N. Transformation of the Forage Grass Caucasian Bluestem via Biolistic Bombardment Mediated DNA Transfer. *PLANT PHYSIOL* 102:167, 1993; Golovkin, M.V., Abraham, M., Morocz, S., Bottka, S., Feher, A., and Dudits, D. Production of Transgenic Maize Plants by Direct DNA Uptake into Embryogenic Protoplasts. *PLANT SCI* 90:41-52, 1993; Guo, G.Q., Xu, Z.H., Wei, Z.M., and Chen, H.M. Transgenic Plants Obtained from Wheat Protoplasts Transformed by Peg Mediated Direct Gene Transfer. *CHIN SCI BULL* 38:2072-2078. 1993; Asano, Y. and Ugaki, M. Transgenic plants of Agrostis alba obtained by electroporationmediated direct gene transfer into protoplasts. *Pl.Cell Rep.* 13, 1994; Ayres, N.M. and Park, W.D. Genetic

Transformation of Rice. *CRIT REV PLANT SCI* 13:219-239, 1994; Barcelo, P., Hagel, C., Becker, D., Martin, A., and Lorz, H. Transgenic Cereal (Tritordeum) Plants Obtained at High Efficiency by Microprojectile Bombardment of Inflorescence Tissue. *PLANT J* 5:583-592, 1994; Becker, D., Brettschneider, R., and Lorz, H. Fertile Transgenic Wheat from Microprojectile Bombardment of Scutellar Tissue. *PLANT J* 5:299-307, 1994; Biswas, G.C.G., Iglesias, V.A., Datta, S.K., and Potrykus, I. Transgenic Indica Rice (Oryza Sativa L) Plants Obtained by Direct Gene Transfer to Protoplasts. *J BIOTECHNOL* 32: 1-10, 1994; Borkowska, M., Kleczkowski, K., Klos, B., Jakubiec, J., and Wielgat, B. Transformation of Diploid Potato with an Agrobacterium Tumefaciens Binary Vector System .1. Methodological Approach. *ACTA PHYSIOL PLANT* 16:225-230, 1994; Brar, G.S., Cohen, B.A., Vick, C.L., and Johnson, G.W. Recovery of Transgenic Peanut (Arachis Hypogaea L) Plants from Elite Cultivars Utilizing Accell(R) Technology. *PLANT J* 5:745-753, 1994; Christou, P. Genetic Engineering of Crop Legumes and Cereals Current Status and Recent Advances. *AGRO FOOD IND HI TECH* 5: 17-27, 1994; Chupeau, M.C., Pautot, V., and Chupeau, Y. Recovery of Transgenic Trees After Electroporation of Poplar Protoplasts. *TRANSGENIC RES* 3: 13-19, 1994; Eapen, S. and George, L. Agrobacterium Tumefaciens Mediated Gene Transfer in Peanut (Arachis Hypogaea L). *Pl.Cell Rep.* 13:582-586, 1994; Hartman, C.L., Lee, L., Day, P.R., and Tumer, N.E. Herbicide Resistant Turfgrass (Agrostis Palustris Huds) by Biolistic Transformation. *BID-TECHNOLOGY* 12:919923, 1994; Howe, G.T., Goldfarb, B., and Strauss, S.H. Agrobacterium Mediated Transformation of Hybrid Poplar Suspension Cultures and Regeneration of Transformed Plants. *Plant Cell Tissue & Orgart Culture* 36:59-71, 1994; Konwar, B.K. Agrobacterium Tumefaciens Mediated Genetic Transformation of Sugar Beet (Beta Vulgaris L). *J PLANTBIOCHEM BIOTECHNOL* 3:37-41, 1994; Ritala, A., Aspegren, K., Kurten, U., Salmenkalliomarttila, M., Mannonen, L., Hannus, R., Kauppinen, V., Teeri, T.H., and Enari, T.M. Fertile Transgenic Barley by

Particle Bombardment of Immature Embryos. *PLANT MOL BIOL* 24:317-325, 1994; Scorza, R., Cordts, J.M., Ramming, D.W., and Emershad, R.L. Transformation of Grape (Vitis Vinifera L) Somatic Embryos and Regeneration of Transgenic Plants. J *CELL BIOCHEM* :102, 1994; Shimamoto, K. Gene Expression in Transgenic Monocots. *CURR OPINBIOTECHNOL* 5:158-162, 1994; Spangenberg, G., Wang, Z.Y., Nagel, J., and Potrykus, I. Protoplast Culture and Generation of Transgenic Plants in Red Fescue (Festuca Rubra L). *PLANT SCI* 97:83-94, 1994; Spangenberg, G., Wang, Z.Y., Nagel, J., and Potrykus, I. Gene Transfer and Regeneration of Transgenic Plants in Forage Grasses. J *CELL BIOCHEM* :102, 1994; Wan, Y.C. and Lemaux, P.G. Generation of Large Numbers of Independently Transformed Fertile Barley Plants. *PLANT PHYSIOL* 104:3748, 1994; Weeks, J.T., Anderson, O.D., and Blechl, A.E. Stable Transformation of Wheat (Triticum Aestivum L) by Microprojectile Bombardment. *J CELL BIOCHEM :104*, 1994; Ye, X.J., Brown, S.K., Scorza, R., Cordts, J., and Sanford, J.C. Genetic Transformation of Peach Tissues by Particle Bombardment. *JAMER SOCHORTSCI* 119:367-373, 1994; Spangenberg, G., Wang, Z.Y., Nagel, J., and Potrykus, I. PROTOPLAST CULTURE AND GENERATION OF TRANSGENIC PLANTS IN RED FESCUE (FESTUCA RUBRA L). *Plant Science 1994* 97:83-94, 1995.

As *nim1* host plants may also be susceptible to pathogens outside of the host-range into which they normally fall, these plants also have significant utility in the molecular, genetic, and biological study of host-pathogen interactions. Furthermore, the UDS phenotype of *nim1* plants also renders them of utility for fungicide screening. *nim1* mutants selected in a particular host have considerable utility for the screening of fungicides using that host and pathogens of the host. The advantage lies in the UDS phenoytpe of the mutant, which circumvents the problems encountered due to hosts being differentially

susceptible to different pathogens and pathotypes, or even resistant to some pathogens or pathotypes.

Pathogens of the invention include but are not limited to viruses or viroids, e.g. tobacco or cucumber mosaic virus, ringspot virus or necrosis virus, pelargonium leaf curl virus, red clover mottle virus, tomato bushy stunt virus, and like viruses; fungi, e.g. *Phythophthora parasitica* and *Peronospora tabacina*; bacteria, e.g. *Pseudomonas syringae* and *Pseudomonas tabaci*; insects such as aphids, e.g. *Myzus persicae*; and lepidoptera, e.g., *Heliothus spp.;* and nematodes, e.g., *Meloidogyne incognita*. The methods of the invention are useful against a number of disease organisms of maize including but not limited to downy mildews such as *Sclerophthora macrospora*, *Sclerophthora rayissiae*, *Sclerospora graminicola*, *Peronosclerospora sorghi*, *Peronosclerospora philippinensis*, *Peronosclerospora sacchari* and *Peronosclerospora maydis*; rusts such as *Puccinia sorphi*, *Puccinia polysora* and *Physopella zeae*; other fungi such as *Cercospora zeae-maydis*, *Colletotrichum graminicola*, *Fusarium monoliforme*, *Gibberella zeae*, *Exserohilum turcicum*, *Kabatiellu zeae* and *Bipolaris maydis*; and bacteria such as *Erwinia stewartii*.

DESCRIPTION OF THE SEQUENCE LISTING

SEQ ID NO:1 - 9919-bp genomic sequence from Figure 14.

SEQ ID NO:2 - 5655-bp genomic sequence from Figure 15.

SEQ ID NO:3 - AA sequence of wild-type NIM protein encoded by cds of seq id no:2.

SEQ ID NO:4 - Rice-1 AA sequence 33-155 from Figure 19.

SEQ ID NO:5 - Rice-1 AA sequence 215-328 from Figure 19.

SEQ ID NO:6 - Rice-2 AA sequence 33-155 from Figure 19.

SEQ ID NO:7 - Rice-2 AA sequence 208-288 from Figure 19.

SEQ ID NO:8 - Rice-3 AA sequence 33-155 from Figure 19.

SEQ ID NO:9 - Rice-3 AA sequence 208-288 from Figure 19.

SEQ ID NO:10 - Rice-4 AA sequence 33-155 from Figure 19.

SEQ ID NO:11 - Rice-4 AA sequence 215-271 from Figure 19.

DEPOSITS

The following vector molecules have been deposited with American Type Culture Collection 12301 Parklawn Drive Rockville, MD 20852 U.S.A. on the dates indicated below:

Plasmid BAC-04 was deposited with ATCC on May 8, 1996 as ATCC 97543.

Plasmid P1-18 was deposited with ATCC on June 13, 1996 as ATCC 97606.

Cosmid D7 was deposited with ATCC on September 25, 1996 as ATC 97736.

EXAMPLES

Example 1

Identifying *NIM1* clones by map-based cloning. High resolution genetic mapping and physical mapping of *NIM1* in Arabidopsis.

1. Plant Material and Isolation of *nim1* Mutants.

Nim1 mutants were isolated from two Arabidopsis ecotype Ws-O plant populations, as described by Delaney et al., (1995) PNAS 92, 6602-6606. One mutant population was in the form of an M2 library derived from ethyl methane sulfonate (EMS) mutagenized seeds (purchased from Lehle, Round Rock, TX) and the other was in the form of a T-DNA population derived from seed obtained from the Ohio State University Arabidopsis Biological Resource Center (Columbus, OH).

The basis of the screen for noninducible immunity (*nim1*) mutants was to screen mutagenized plant populations for plants in which resistance to a virulent pathogen could not be induced by INA (2,6 dichloro isonicotinic acid; Metraux, et al., 1991. In: Advances in Molecular Genetics of Plant-Microbe Interactions. Vol 1, 432-439. Hennecke and Verma, eds.; Kessmann et al. 1993 In: Mode of action of agrochemicals. Y Honma, ed.; Vernooij et al, 1995, Molec. Pl. Microbe Interaction 8, 228-234).

Plants from the mutant populations were grown at high density in large trays in commercial planting mix. When the plants were 2 weeks of age, the trays were sprayed with 0.25 mg/ml INA. Four days later, the plants were sprayed with a spore suspension of *Peronospora parasitica*, isolate EmWa (EmWa), at $5x10^4$ to $1x10^5$ spores/ml. This fungus is normally virulent on the Arabidopsis Ws-O ecotype, unless resistance is first induced in these plants with INA or a similar compound.

Following incubation in a high humidity environment, plants with visible disease symptoms were identified, typically 7 days after the infection. These plants did not show resistance to the fungus, despite the application of the resistance-inducing chemical and were thus potential *nim* (non-immunity) mutant plants. From 360,000 plants, 75 potential *nim* mutants were identified.

These potential mutant plants were isolated from the flat, placed under low humidity conditions and allowed to set seed. Plants derived from this seed were screened in an identical manner for susceptibility to the fungus EmWa, again after pretreatment with INA. The progeny plants that showed infection symptoms were defined as *nim* mutants. Six *nim* lines were thus identified. One line (*nim1*) was isolated from the T-DNA population and five from the EMS population.

2. Scoring Plant Reactions to INA and Other Chemical Inducers of Disease Resistance.

i. Phenotypic analysis of *nim1*.

Salicylic acid (SA) and benzo(1,2,3)thiadiazole-7-carbothioic acid S-methyl ester (BTH) are two chemicals that, like INA, induce broad spectrum disease resistance, termed Systemic Acquired Resistance (SAR), in wildtype plants. Since INA did not induce resistance in the *nim1* plants, these plants were also evaluated for their disease resistance response following pretreatment with SA and BTH (as partly described in Delaney et al, 1995, PNAS 92, 6602-6606).

Plants were sprayed with 1, 5, or 15 mM SA or 0.25 mg/ml BTH and challenge inoculated with EmWa 5 days later (as described in example 1 above). Both SA and BTH failed to protect *nim1* plants from fungal infection, as evidenced by the presence of disease symptoms and fungal growth on these plants. Thus, the *nim1* plants were not responsive to any of the SAR-inducing chemicals, implying that the mutation was downstream of the entry point(s) for these chemicals in the resistance induction pathway.

Nim1 was also evaluated for its disease response to infection with 2 incompatible *P. parasitica* isolates, Wela and Noco (i.e. these fungal strains do not cause disease on wildtype Ws-O plants). *nim1* plants were sprayed with conidial suspensions of 5-10x10^4 spores/ml of Wela or Noco and incubated under high humidity for 7 days. Unlike wildtype plants, *nim1* plants developed disease symptoms in response to both Wela and Noco infection. The symptoms were necrotic flecking and trailing, with some sporulation. Following lactophenol blue staining, fungal hyphae were easily observed in the leaves of *nim1* plants. Thus, the *nim1* plants are susceptible to normally incompatible *P. parasitica* isolates. This result shows that the *nim1* plants are not only defective in chemically induced

disease resistance, but are also defective in natural resistance to microorganisms that are normally not pathogenic.

ii. Biochemical analysis of *nim1*.

SA, INA and BTH induce SAR and expression of the SAR genes, which include the Pathogenesis Related genes PR-1, PR-2 and PR-5 in *Arabidopsis*. Since these compounds did not induce disease resistance in *nim1* (as described in example 1.2 above), this mutant line was analyzed for SAR gene expression following SA, INA or BTH treatment.

After treatment of *nim1* plants with SA, INA or BTH, plant tissue was harvested and analyzed for accumulation of RNA from the PR-1, PR-2 and PR-5 genes. To this end, total RNA was isolated from the treated tissues and electrophoresed on an agarose gel. Triplicate gel blots were prepared and each was hybridized with a probe for one of these 3 SAR genes as described in Delaney et al, 1995, PNAS 92, 6602-6606. In contrast to the case in wildtype plants, the chemicals did not induce RNA accumulation from any of these 3 SAR genes in *nim1* plants, as shown in figure 1. Taken together, the results indicate that the chemicals induce neither SAR nor SAR gene expression in *nim1* plants.

Since the chemicals did not induce SAR, or SAR gene expression in *nim1* plants, it was of interest to investigate whether pathogen infection could induce SAR gene expression in these plants, as it does in wildtype plants. Ws-O and *nim1* plants were sprayed with EmWa spores as described and tissue collected for RNA analysis at several timepoints. Pathogen infection (EmWa) of wildtype Ws-O plants induced PR-1 gene expression within 4 days after infection, as shown in figure 2. In *nim1* plants, however, PR-1 gene expression is not induced until 6 days after infection and the level is reduced relative to the wildtype at that time. Thus, following pathogen infection, PR-1 gene expression in *nim1* plants is delayed and reduced relative to the wildtype.

Infection of wildtype plants with pathogens that cause a necrotic reaction leads to accumulation of SA in the infected tissues. It has been shown that this endogenous SA is required for signal transduction in the SAR pathway, i.e., breakdown of the endogenous SA leads to a decrease in disease resistance. This defines SA accumulation as a marker in the SAR pathway (Gaffney et al, 1993, Science 261, 754-756).

Nim1 plants were tested for their ability to accumulate SA following pathogen infection. *Pseudomonas syringae* tomato strain DC 3000, carrying the avrRpt2 gene, was injected into leaves of 4-week-old *nim1* plants. The leaves were harvested 2 days later for SA analysis as described by Delaney et al, 1995, PNAS 92, 6602-6606. This analysis showed that the *nim1* plants accumulated high levels of SA in infected leaves, as shown in Figure 3. Uninfected leaves also accumulated SA, but not to the same levels as the infected leaves, similar to what has been observed in wild-type *Arabidopsis*. This indicated that the *nim* mutation maps downstream of the SA marker in the signal transduction pathway. This was anticipated, as INA and BTH (inactive in *nim1* plants) are known to stimulate a component in the SAR pathway downstream of SA (Vernooij et al., 1995, Molec. Pl. Microbe Interaction 8, 228-234; Friedrich, et al., 1996, The Plant Journal 9, in press; and Lawton, et al., 1996, The Plant Journal 9, in press). In addition, as described in Example 1.2, exogenously applied SA did not protect *nim1* from EmWa infection.

3. Genetic analysis of *nim1*.

Nim1 plants were backcrossed to wild-type Ws-O plants, and F1 progeny were tested for resistance to EmWa after INA pretreatment, as described in Example 1.1 above. None of the INA-pretreated F1 plants had infection symptoms, whereas the *nim1* control plants did show infection. Hence, the *nim1* mutation was determined to be recessive.

The F2 population from the Ws-O x *nim1* cross was also assayed for its disease resistance after INA pretreatment. Of this population, approximately 1/4 (32/130 plants) showed disease symptoms after EmWa treatment of INA pre-sprayed plants and 3/4 (98/130 plants) showed no disease. These results indicate that the *nim* mutation identifies a single genetic locus and corroborates the F1 data that show the recessive nature of the mutation.

4. Identification of markers in and genetic mapping of the *NIM* locus.

For conventional map-based cloning of the *NIM* gene, markers had to be identified that were genetically closely linked to the mutation. This was accomplished in 2 steps. First, the *nim1* plants were crossed to a different Arabidopsis genotype, Landsberg erecta (Ler), and F2 plants from this cross which had a *nim1* phenotype (i.e. plants that are homozygous nim/nim at the *NIM* locus) were identified. From these, plants that had a Ler genotype at a nearby DNA marker were identified by molecular analysis. These plants, by virtue of the identification criterion, are recombinant between the marker and the *NIM* locus. The frequency of recombinants defines the genetic distance between the marker and the *NIM* locus.

The second prerequisite for map based cloning is that markers are identified that are genetically very close to the *NIM* locus, i.e. markers that identify very few recombinants. If genetic markers are identified that are very close, than these can be used to isolate genomic DNA clones that are close to the *NIM* locus. The *NIM* locus can then be cloned by walking, if not already present on the cloned DNA. Walking can be initiated from both sides of the gene. It relies on obtaining overlapping clones that are successively closer to the gene of interest. When a single DNA marker is obtained from a walk initiated from, for instance, the North end and it identifies no recombinants between this marker and the gene of interest, it must be very close to the gene. However, if the marker does identify

recombinant(s) from the South end, the clone from which the marker was obtained must have crossed the gene. By definition then, the gene of interest is cloned. It must be located between this marker and the last North-end marker that identifies the least number of recombinants from the North end.

In a first step, a large number of recombinants are generated by genetic crossing. In a second step, recombinants that are close to the *NIM* gene are identified with the use of molecular markers. Many markers have been described in the literature and several methods exist to develop additional markers. Our approach has relied on a number of marker systems, including SSLPs and AFLPs (see below).

i. Genetic crosses.

In order to map the chromosomal position of the *NIM* gene relative to the SSLP and AFLP markers, *nim1* was crossed to Ler to make a mapping population. F2 plants from this cross were grown and leaves harvested for future DNA extractions. Next, the F2 plants were scored for the *nim1* phenotype, as described in example 1.1 above. Also, F3 populations derived from individual F2 plants were grown and scored for the *nim* phenotype. DNA was extracted from the stored tissue of *nim1* phenotype F2 and F3 plants by the CTab method, as described (Rogers and Bendich, 1988, Plant Molecular Biology Manual, A6, 1-10). This DNA was used for mapping the *NIM* gene, as described below.

ii. Simple Sequence Length Polymorphism markers.

The Simple Sequence Length Polymorphism (SSLP) markers ATHGENEA and nga111 have been described (Bell and Ecker, 1994, Genomics 19, 137-144). Primers used for detection of these SSLPs are listed in Table 1.

Table 1. SSLP primer sequences.

primer set	primer sequence (5' to 3')
ATHGENEA (1)	ACC ATG CAT AGC TTA AAC TTC TTG
	ACA TAA CCA CAA ATA GGG GTG C
ATHGENEA (2)	ACC ATG CAT AGC TTA AAC TTC TTG
	CCA AAT GTC AAA ATA CTC GTC
nga111 (1)	CTC CAG TTG GAA GCT AAA GGG
	TGT TTT TTA GGA CAA ATG GCG
nga111 (2)	CTC CAG TTG GAA GCT AAA G
	TGT TTT TTA GGA CAA ATG G

Genetic mapping of the *NIM* gene relative to marker ATHGENEA.

Using the ATHGENEA (1) primers for PCR amplification of Ler genomic DNA, a 205-basepair (bp) band was expected, whereas with Ws-O genomic DNA a band of 211 bp was expected (Bell and Ecker, 1994, Genomics 19, 137-144). The amplification products proved to be difficult to separate on conventional agarose gels. Hence, two alternative methods were developed for separation and detection of these PCR fragments.

In a first method, primerset ATHGENEA (1) (Table 1) was used to amplify genomic DNA in the presence of 6-carboxyrhodamine-labelled UTP (dUTP-R110, obtained from ABI), yielding rhodamine-labelled PCR fragments. The PCR reactions were analyzed on a DNA Sequencer, which detects DNA fragments with single nucleotide resolution.

The specific reagents were: 1xPCR buffer, 2 mM MgCl2, dNTPs each 200 mM, 2 mM dUTP-R110, ATHGENEA (1) primers at 0.75 mM, 10 ng DNA and 0.75 units Taq polymerase in a 20 ml reaction volume. Amplification conditions were: 3 minutes 94°C followed by 35 cycles of 15 seconds at 94°C, 15 seconds at 55°C and 30 seconds at 72°C. These samples were analyzed on an ABI 377 DNA Sequencer, capable of detecting fluorescently labeled DNA fragments with single nucleotide (nt.) resolution. This allowed for genotyping the plant samples: a 205-nucleotide DNA fragment was obtained from Ler DNA and a 211-nucleotide band from Ws-O DNA. Thus, DNA fragments differing by 6 nucleotides in length could be easily distinguished, allowing for easy genotyping of samples as homozygous Ws-O, homozygous Ler and heterozygous Ws-O/Ler at the ATHGENEA locus.

In order to increase the throughput of this system, a multiplexing scheme was used. Some DNA samples were PCR amplified as described above with primer set ATHGENEA (1) , whereas other samples were analyzed with primerset ATHGENEA (2) (listed in table 2), in each case in the presence of 6-carboxyrhodamine-labeled dUTP. Primer set ATHGENEA (2) was made based on the published sequence of ATHGENEA (Simoens et al., 1988, Gene 67, 1-11). This primerset amplified a DNA fragment of 139 bp from Ler DNA and a 145-bp band from Ws-O DNA. Amplification reaction conditions for primerset ATHGENEA (2) were identical to those described for primerset ATHGENEA (1), above.

Single reactions using primer set ATHGENEA (1) and single reactions using primer set ATHGENEA (2) were mixed together before electrophoresis on the ABI 377 DNA Sequencer. This multiplexing approach allowed for genotyping 2 samples in a singe lane of the Sequencer, one at positions 145/139 nt. and one at positions 211/205 nt. on the Sequencer.

In the second method, PCR fragments were labelled by using a primer labelled with the fluorescent dye FAM-6 (6-carboxyfluorescein) (Integrated DNA Technologies, Inc.). The forward ATHGENEA primers of the ATHGENEA (1) and (2) primer sets are identical in sequence (see Table 1). This primer was labeled with FAM-6 and used in a PCR amplification reaction with the following reagents (Perkin Elmer): 1xXL buffer, 1 mM MgCl2, dNTPs each at 200 mM, primers each at 0.50 mM (forward primer FAM-6 labeled), 10 ng genomic DNA and 0.5 units XL polymerase in a 20-ml reaction volume. The cycling conditions were: 3 minutes at 94°C, followed by 35 cycles of 15 seconds at 94°C, 15 seconds at 59°C and 30 seconds at 72°C. Again, single reactions using primer set ATHGENEA (1) and single reactions using primer set ATHGENEA (2) were mixed together before electrophoresis on the ABI 377 DNA Sequencher. This multiplexing approach allowed for genotyping 2 samples in a single lane of the Sequencer, one at positions 145/139 nt. and one at positions 211/205 nt.

All F2 and F3 samples from *nim1* phenotype plants were scored for their genotype at the ATHGENEA locus as described above. All samples that were heterozygous at this locus identified plants that were recombinant between the *NIM1* locus and the ATHGENEA locus. In a population of 1144 F2 *nim1* phenotype plants and F3 *nim1* phenotype populations that were scored in this way, 98 were heterozygous at the ATHGENEA locus, giving an estimate of the genetic distance between this SSLP locus and the *NIM1* locus of 4.3 cM. This established that the *NIM1* locus was on chromosome 1, near the ATHGENEA marker.

Genetic mapping of the *NIM1* gene relative to marker nga111.

Two primer sets for SSLP marker nga111 (described in Bell and Ecker, 1994, Genomics 19, 137-144) were used to amplify genomic DNA of F2 and F3 *nim1* phenotype

plants and control Ws-O and Ler plants. Primer set nga111 (1) (described in Bell and Ecker, 1994, Genomics 19, 137-144 and listed in Table 1) was used under the following conditions: 1xPCR buffer, 2 mM MgCl2, dNTPs each at 200 mM, primers at 0.75 mM, 10 ng DNA and 0.75 unit Taq polymerase in a 20 ml reaction volume. Primer set nga111 (2) (listed in Table 1, and a derivative of primer set nga111 (1)) was used under different conditions: 1xPCR buffer, 1.5 mM MgCl2, dNTPs each 200 mM, primers at 1 mM, 10 ng DNA and 1 unit Taq polymerase in a 20 ml reaction volume. Both reactions were amplified by incubation at 94°C for 1 minute, followed by 40 cycles of 15 seconds at 94°C, 15 seconds at 55°C and 30 seconds at 72°C.

The samples were analyzed on 3-5% agarose gels. The band obtained from amplification of Ws-O DNA with either primer set was 146 bp, whereas amplifying Ler DNA resulted in a 162-bp band. Plant samples that were heterozygous at the nga111 locus identified plants that were recombinant between this SSLP marker and the *NIM* locus. Among 1144 F2 *nim1* phenotype plants and F3 *nim1* phenotype populations, 239 were identified as heterozygous for the nga111 marker, giving an estimate for the genetic distance between the SSLP marker and the *NIM* locus of 10.4 cM. This corroborated that the *NIM1* locus was on chromosome 1. Since few *nim1* phenotype plants existed that were heterozygous at both ATHGENEA and nga111, the *NIM1* locus was determined to be between these 2 markers, with ATHGENEA located North of the *NIM1* gene and nga111 located South of the *NIM1* gene. This placed the *NIM1* gene approximately 10 cM north of nga111, near position 85 on chromosome 1 (Lister and Dean, 1993, Plant J. 4, 745-750; Bell and Ecker, 1994, Genomics 19, 137-144).

iii. Amplified Fragment Length Polymorphism markers.

For map-based cloning of the *NIM1* gene, it is necessary to identify molecular markers that are successively closer to this gene. For this purpose, Amplified Fragment Length Polymorphism (AFLP) markers were generated by using the selective restriction fragment amplification method described by Zabeau and Vos (1993, European Patent Application EP 0534858) and Vos et al. (1995, Nucleic Acid Research 23, 4407-4414).

Outline of the AFLP Technology.

The use of the AFLP technology in mapping relies on selective amplification of a set of DNA bands in 2 genetically distinct samples. Finding that any of the obtained bands are different between the 2 genotypes identifies those bands as markers for that genotype. If the marker cosegregates at high frequency with the gene (mutation) of interest, then the marker is close to the genetic locus.

Selective amplification of a small set of DNA fragments in a complex DNA sample is achieved in a 2-step process. First, DNA fragments are generated by digesting the DNA with restriction enzymes, followed by ligation of adapters to the ends. Second, primers consisting of a sequence complementary to the adapters plus a 3' extension (typically 0-3 nucleotides) are used to amplify only those DNA fragments with ends that are complementary to these primers. If a single nucleotide extension is used, then theoretically, each primer will "fit" on approximately 1/4 of all fragments, with 1/16 of all fragments having a primer fit on both ends. Thus, a limited set of DNA fragments is amplified with these primers. By further radiolabelling one primer, an even smaller subset of visible bands can be obtained.

AFLP analysis.

For AFLP analysis of DNA samples, 50 ng DNA was digested with the appropriate enzymes (usually EcoRI and MseI; see below) and adapters (listed in table 2 below) were ligated to the restriction fragments (usually EcoRI and MseI). The sequences of the primers and the YAC, P1 and BAC clones are described in detail below. The templates were used for amplification reactions (approximately 0.5 ng DNA per reaction), using primers that were complementary to the adapters, with short 3' extensions (2 or 3 nucleotides; primer sequences are listed below). Since one of the primers is radioactively labelled (usually the EcoRI primer), only a subset of the amplified fragments is visible upon autoradiographic analysis of the gel used to separate the bands.

Amplification conditions for cloned DNA (YAC, P1, cosmid) were as follows: 36 cycles of 30 sec. at 94°C (denaturation), 30 sec annealing and 1 min extension at 72°C. The annealing temperature in the first cycle was 65°C and was reduced by 0.7°C in each cycle for the next 12 cycles and then kept at 56°C. For genomic DNA of *Arabidopsis* plants, the amplification was performed in 2 steps: in the first step (preamplification), the DNA was amplified with primers that have a single nucleotide extension (neither primer was labeled). Reaction conditions for this amplification reaction were: 20 cycles of 30 sec. denaturation (94°C), 1 min annealing (56°C) and 1 min extension (72°C). In the second step, the first amplification reaction was diluted 10 fold and reamplified 36 cycles with primers containing the full-size extensions (using one labeled primer) under the following conditions: 30 sec. at 94°C (denaturation), 30 sec annealing and 1 min extension at 72°C. The annealing temperature in the first cycle was 65°C and was reduced by 0.7°C in each cycle for the next 12 cycles and then kept at 56°C. The final reaction products were separated on a polyacrylamide gel and the gel was exposed to film, allowing visualization of the radiolabeled PCR bands. When this procedure was applied to DNA from 2 genotypes simultaneously, AFLP bands were identified that were diagnostic for one genotype or the other. Such bands are called informative AFLP bands, or AFLP markers. Table 2 shows Adapters used in the AFLP analysis.

Table 2

enzyme	adapter
EcoRI	5'-CTCGTAGACTGCGTACC-3'
	3'-CATCTGACGCATGGTTAA-5'
HindIII	5'-CTCGTAGACTGCGTACC-3'
	3'-CATCTGACGCATGGTCGA-5'
PstI	5'-CTCGTAGACTGCGTACATGCA-3'
	3'-CATCTGACGCATGT-5'
MseI	5'-GACGATGAGTCCTGAG-3'
	3'-TACTCAGGACTCAT-5'

Generation of AFLP markers and fine mapping of the *NIM1* locus.

A population of recombinant inbred lines derived from a cross between the *Arabidopsis* ecotypes Landsberg erecta (Ler) and Columbia (Col) (Lister and Dean, 1993, Plant J. 4, 745-750) was used for AFLP marker screening. The primers used for the AFLP screening were:

EcoRI-primers: 5'-GACTGCGTACCAATTCWN-3'

MseI-primers: 5'-GATGAGTCCTGAGTAAXWN-3'

An "N" in the primers indicates that this part was variable (A, C, G or T), a "W" indicates A or T, and an "X" indicates a C. All 8 possible primers were used for both the

EcoRI- and MseI-primer. This gave a total of 64 (8 x 8) primer combinations (PCs) that were used to amplify DNA from the recombinant inbred line and the parental genotypes, Ler and Col, as described above. The amplification reactions were run on a denaturing polyacrylamide gel to separate AFLP fragments by size and the gel was exposed to film. The film was inspected for bands that were present in only one genotype, i.e. inspected for AFLP markers.

The AFLP markers, i.e., DNA fragments that are polymorphic between both parents of the recombinant inbred lines, were used for constructing a genetic map of the recombinant inbred line population. Example 1.5i below describes the mapping of the *NIM1* gene on *Arabidopsis* chromosome 1, at approximately position 85. Those AFLP markers that had been mapped (using the recombinant inbred line) between positions 81 and 88 of Arabidopsis chromosome 1 were chosen for analyzing recombinant plants for the presence of said AFLP markers and thus for mapping the *NIM1* gene more precisely. Seven AFLP markers from this region were identified as being informative; they were polymorphic between both parents of the *nim1*xLer cross. Six AFLP markers were Ler-specific, i.e. these AFLP markers were absent in Ws (and in Col as well). One AFLP marker was Ws-specific, i.e. a Col-specific AFLP marker (absent in Ler) was also present in Ws. These AFLP markers are: L81.1, L81.2, W83.1, L84, L85, L87 and L88 (an L-marker is specific for ecotype Ler and a W-marker is specific for both the ecotypes Col and Ws; the number indicates the map position). These AFLP markers were used to analyze recombinant plants from the *nim1*xLer cross (see below). In addition, AFLP marker C86 (a recombinant, inbred-line-derived marker specific for Col) was used in isolating DNA clones (see below). Table 3 lists the primer sequences that were used to obtain these AFLP markers.

Table 3 shows primer combinations of AFLP markers derived from recombinant inbred line population.

"EcoRI-" refers to the sequence 5'-GACTGCGTACCAATTC-3' and

"MseI-" refers to the sequence 5'-GATGAGTCCTGAGTAA-3'.

Table 3

AFLP marker	Corresponding primer combinations	
L81.1	EcoRI-CA	MseI-CCG
L81.2	EcoRI-AA	MseI-CAA
W83.1	EcoRI-CA	MseI-CTC
L84	EcoRI-AAT	MseI-CAA
L85	EcoRI-CA	MseI-CCT
L87	EcoRI-CA	MseI-CTT
L88	EcoRI-AG	MseI-CTA
C86	EcoRI-AG	MseI-CCT

A detailed genetic map of the region was constructed using the AFLP markers described above by typing the recombinants. A total of 337 recombinant plants were available out of 1144 F2 *nim1* plants. These recombinants were first screened with the North-flanking AFLP markers L81.2 and ATHGENEA and the South-flanking markers L88 and nga111. Forty-eight plants were homozygous *nim1/nim1* and heterozygous at ATHGENEA and L81.2, and 21 plants were homozygous *nim1/nim1* and heterozygous at nga111 and L88. These recombinant plants were further analyzed with 9 AFLP markers in the *NIM* region, including 4 AFLP markers that were derived from the recombinant inbred line mapping population, (W83.1, L84, L85 and L87) and 5 AFLP markers derived from analysis of YAC clones (W83.3/W84.1, W84.2, W85.1, W86.1 and L86, see below).

The genetic map of *NIM1*, based on this analysis, is depicted in Figure 4. As seen, 27 recombinants were found between marker W84.2 and *NIM1* and 14 recombinants were

found between W85.1 and *NIM1*. Marker L85 is linked closely to *NIM1*, but this marker could not be mapped on the YAC, BAC or P1 clones (see below) and was, therefore, not useable for identification of the *NIM1* gene.

5. Physical mapping of the *NIM1* region.

i. Isolation of YAC clones using AFLP markers closely linked to *NIM1*.

The CIC library, an Arabidopsis ecotype Columbia YAC library (Bouchez et al, 1995, 6th Int. Conf on Arabidopsis Research, Madison, WI), was screened for YAC clones in the *NIM* region. This library has about 2.5 nuclear genome equivalents and has an average insert size of 450 kb. The YAC library was screened with two AFLP markers: W83.1 and C86. W83.1 is the most closely linked recombinant, inbred-line-derived AFLP marker north of *NIM1*, and C86 is a recombinant, inbred-line-derived AFLP marker specific for Col (absent in Ler and Ws). C86 mapped south of the *NIM1* gene on the map of the recombinant inbred line population. This Col AFLP marker has been used instead of the closely linked Ler AFLP markers (Figure 4), because the latter AFLP markers detected only ecotype *Landsberg erecta* and hence cannot be used for screening the Columbia YAC library.

The YAC library was screened in two steps. Firstly, the cells of the YAC clones of each plate of the twelve 96-well microtiter plates were pooled (a plate pool) and used for DNA isolation as described by Ross et al (1991, Nucleic Acids Res. 19, 6053). The pools were screened with both AFLP markers. Subsequently, from each positive plate pool, the DNA samples of each row (a pool of 8 clones) and of each column (a pool of 12 clones) were screened with the AFLP marker for which the plate pool was positive. In this way, the individual positive YAC clones could be identified. The screening yielded a total of 4 YAC clones: YAC 12F04 and YAC 12H07 were isolated using the North AFLP marker W83.1,

and YAC 10G07 and YAC 7E03 using the south AFLP marker C86 (for the nomenclature of the YAC clones the CIC numbering is used). The YACs were "fingerprinted" by AFLP, giving YAC-specific AFLP fragments. Fingerprints of the YACs were compared and used to estimate overlaps between the YACs (see also Tables 5 and 6). Based on the AFLP fingerprints, clone 7E03 is essentially covered by clone 10G07 (see also table 5) and clone 12H07 is likewise essentially covered by clone 12F04 (see also table 6).

ii. Generation of AFLP markers from YAC clones.

Since the AFLP markers described above were genetically relatively far from the *NIM1* gene (see figure 3), additional AFLP markers were developed in an effort to find markers that were closer to the *NIM* gene.

A screening for additional YAC-derived AFLP markers was performed on DNA samples of the following: DNA of the isolated YAC clones (4 YACs were identified, as described above), the yeast strain without a YAC, and the three Arabidopsis ecotypes Col, Ler and Ws. In this way, the fragments specific for the YAC clones (absent in the yeast strain and present in Col) could be tested for polymorphism in Ler and Ws (the parents of the recombinant plants identified in Example 1.5 below). All identified polymorphic fragments would thus be additional AFLP markers. In the first AFLP screening the enzyme combination (EC) EcoRI/MseI was used. In this screening two YAC clones, 10G07 and 7E03 (detected with AFLP marker C86, see below), the yeast strain without a YAC and the three Arabidopsis ecotypes Col, Ler and Ws were assayed. The primer combinations with the selective extensions used can be divided into three groups and are depicted in Table 4. A total of 256 (64 + 96 + 96) primer combinations were screened.

In Table 4 below the primer sequences used in the AFLP screening of two YAC clones, 10G07 and 7E03, the yeast strain without a YAC, and the three Arabidopsis

ecotypes Col, Ler and Ws are shown. Three groups of primer combinations have been used. An "N" in the primers indicates that this part was variable (A, C, G or T), an "S" indicates C or G, a "W" indicates A or T, and a "Y" indicates C or T.

Table 4

EcoRI-primers:

5'-GACTGCGTACCAATTCGW-3'

5'-GACTGCGTACCAATTCTS-3'

MseI-primers:

5'-GATGAGTCCTGAGTAAAAS-3'

5'-GATGAGTCCTGAGTAAASA-3'

5'-GATGAGTCCTGAGTAAATN-3'

5'-GATGAGTCCTGAGTAACAN-3'

5'-GATGAGTCCTGAGTAACTN-3'

EcoRI-primers:

5'-GACTGCGTACCAATTCAN-3'

5'-GACTGCGTACCAATTCCW-3'

5'-GACTGCGTACCAATTCTW-3'

MseI-primers:

5'-GATGAGTCCTGAGTAAAAS-3'

5'-GATGAGTCCTGAGTAAASA-3'

5'-GATGAGTCCTGAGTAAGAY-3'

5'-GATGAGTCCTGAGTAAGTW-3'

5'-GATGAGTCCTGAGTAATCG-3'

5'-GATGAGTCCTGAGTAATCT-3'

5'-GATGAGTCCTGAGTAATGW-3'

EcoRI-primers:

5'-GACTGCGTACCAATTCGW-3'

5'-GACTGCGTACCAATTCTN-3'

MseI-primers:

5'-GATGAGTCCTGAGTAAGAW-3'

5'-GATGAGTCCTGAGTAAGCW-3'

5'-GATGAGTCCTGAGTAAGTW-3'

5'-GATGAGTCCTGAGTAATAN-3'

5'-GATGAGTCCTGAGTAATCW-3'

5'-GATGAGTCCTGAGTAATGW-3'

5'-GATGAGTCCTGAGTAATTS-3'

In total, 83 Col-specific fragments were generated, of which 62 were shared by both YAC clones. Three fragments were AFLP markers polymorphic between Ws and Ler, of which two were Ws AFLP markers (a Col fragment also present in Ws and absent in Ler) and one was a Ler AFLP marker (a Col fragment also present in Ler and absent in Ws). These results are presented in Table 5 below.

Table 5 shows a number of shared and unique AFLP fragments detected in YACs 10G07 and 7E03 and the number of informative AFLP markers among these fragments in Ws and Ler genotypes.

Table 5

	AFLP fragments in YAC clones		AFLP marker	
	10G07	7E03	Ws	Ler
shared	62	62	2	1
unique	21	0	0	0

This AFLP analysis thus yielded 3 new AFLP markers (see Figure 4 and below). Their positions relative to each other and relative to the recombinant, inbred-line-derived AFLP markers were determined by analysis of the recombinants with these AFLP markers.

A second screening for AFLP markers was performed assaying all four identified YAC clones (see below) and using the enzyme combination PstI/MseI. The primers used are:

PstI-primers:

5'-GACTGCGTACATGCAGAN-3'

5'-GACTGCGTACATGCAGCW-3'

5'-GACTGCGTACATGCAGGW-3'

5'-GACTGCGTACATGCAGTN-3'

MseI-primers:

5'-GATGAGTCCTGAGTAAAN-3'

5'-GATGAGTCCTGAGTAACW-3'

5'-GATGAGTCCTGAGTAAGW-3'

5'-GATGAGTCCTGAGTAATN-3'

An "N" in the primers indicates that this part was variable (A, C, G or T) and a "W" in the primers indicates that this was A or T. A total of 144 (12 x 12) primer combinations was

screened on all four isolated YAC clones, 12F04, 12H07, 10G07 and 7E03; the yeast strain without a YAC; and the three Arabidopsis ecotypes Col, Ler and Ws. In total, 219 AFLP fragments were generated, of which 144 were present in YAC clones 12F04 and 12H07 (72 were unique for clone 12F04 and 72 were shared between both YACs) and of which 75 were present in YAC clones 10G07 and 7E03 (33 were unique for clone 10G07 and 42 were shared between the 2 YACs). Three fragments derived from the first set of YAC clones were polymorphic (Ws AFLP markers). These results are presented in Table 6 below.

Table 6 lists the number of shared and unique AFLP fragments detected in YACs and the number of informative AFLP markers among these fragments in Ws and Ler genotypes.

Table 6

	number of AFLP fragments in YAC clones				AFLP markers	
	12F04	12H07	10G07	7E03	Ws	Ler
shared	72	72	0	0	1	0
unique	72	0	0	0	2	0
shared	0	0	42	42	0	0
unique	0	0	33	0	0	0

The results indicate that YAC clone 12H07 is part of the larger YAC clone 12F04, and that YAC clone 7E03 is a part of the larger YAC clone 10G07. These data indicate that the larger YAC clones, 12F04 and 10G07, do not overlap. These data do not allow the positioning of the *NIM1* gene on any of these YAC clones. The whole screening, involving 400 primer combinations producing 302 AFLP fragments in the *NIM* region, yielded 5 useful AFLP markers, of which 4 were Ws-specific and one Ler-specific. These 5 additional AFLP

markers have been mapped by analysis of recombinant plants (see figure 4 and below) and are denominated W84.1 (a.k.a. W83.3), W84.2, W85.1, W86.1 and L86.

Table 7 lists the primer sequences used to obtain these AFLP markers. These 5 additional AFLP markers raised the total number of AFLP markers to 12 in the region from L81.1 to L88 (see figure 4 and below).

Table 7 shows primer combinations of AFLP markers derived from YAC clones.
"EcoRI-" refers to the sequence 5'-GACTGCGTACCAATTC-3',
"MseI-" refers to the sequence 5'-GATGAGTCCTGAGTAA-3' and
"PstI-" refers to the sequence 5'-GACTGCGTACATGCAG-3'.

Table 7

AFLP marker	Primer combination with selective extensions	
W84.1	PstI-AT	MseI-TT
W84.2	PstI-AA	MseI-TT
W85.1	EcoRI-CT	MseI-GTA
W86.1	EcoRI-GT	MseI-CTT
L86	EcoRI-GT	MseI-CTT

This information was used to construct a physical map of the region, as shown in Figure 5, with approximate positions of the YAC clones, relative to the genetic map. The map showed that the region containing the *NIM1* locus, between markers W83.1 and W85.1, is partly covered by 3 YAC clones: 12F04 and 10G07/7E03.

iii. Construction of a P1/BAC contig containing the *NIM1* gene.

In the previous sections it was described how AFLP markers linked to the *NIM1* region were isolated and how YACs corresponding to these markers were identified and mapped. The results obtained while localizing the *NIM1* gene to a chromosome fragment, did not allow the definition of a specific DNA segment containing the *NIM1* gene: the flanking AFLP markers were mapped to different YACs that did not overlap. It was, therefore, not possible to determine the precise physical position of the *NIM1* gene; it could be located on either of the two YACs or in the gap between the YACs. An alternative approach was selected to close the physical gap between the flanking markers: a P1 and BAC library were employed to bridge the gap between the flanking AFLP markers.

The libraries used for gap closure were an Arabidopsis ecotype Columbia P1 library described by Liu et al (The Plant J. 7, 351-358, 1995) and an ecotype Columbia BAC library described by Choi et al (http/genome-www.stanford.edu/Arabidopsis/ww/Vol2/ choi.html). The P1-library consists of about 10,000 clones with an average insert size of 80 kb and the BAC library consists of about 4000 clones with an average insert size of 100 kb. In theory these libraries represent about 10 nuclear genome equivalents (assuming a haploid genome size for *Arabidopsis* of 120 Mb).

iv. Identification of P1 clones corresponding to the flanking markers.

The flanking markers Ws84.2 and Ws85.1 were used to screen pools of P1 clones using a similar strategy as previously described for screening of the YAC library (see Example 1.5i). P1 clones having the marker fragments were selected and "plasmid" DNA was isolated. The various P1 clone DNAs were fingerprinted using the ECs EcoRI/MseI and HindIII/MseI and primers without selective nucleotides. A physical map was constructed, i.e. a map giving the size and overlaps of the clones, by comparing the AFLP fingerprints. The number of AFLP fragments that are unique and the number of AFLP fragments that are common between clones indicate the extent of the overlaps. The map is displayed in Figure 6. The AFLP fingerprinting revealed that two sets of non-overlapping P1-contigs had been constructed each containing one of the flanking markers: P1-1 and P1-2 containing marker Ws84.2; P1-3 and P1-4 containing marker W85.1. Consequently, the gap between the flanking markers was not closed (Figure 6).

The positions of the P1 contigs with respect to the YAC contig was determined by AFLP fingerprinting of the YACs and P1 clones with a number of YAC-specific PCs described above. P1 clones P1-1 and P1-2 appeared to overlap completely with YAC CIC12F04, but only partially with YAC CIC12H07. Therefore, the latter P1 clones could be positioned on the YAC contig CIC12H07/12F04 (Figure 6). P1 clones P1-3 and P1-4 overlapped completely with both YACs CIC7E03 and CIC10G07 and it appeared that AFLP marker W86.1, like W85.1, was mapped to this P1 contig (Figure 6).

Next, marker L85 was used to identify corresponding P1 and BAC clones. L85 is an ecotype-Landsberg-specific marker and, therefore, colony hybridization of radioactively labelled L85 DNA to P1 and BAC filters was employed. Not a single P1 or BAC clone hybridizing to L85 was identified. This supported our earlier findings that the L85 sequence

is lacking in the *Arabidopsis* ecotype Columbia genome and is, therefore, the most likely explanation of why no corresponding clones were identified.

v. Extending the *NIM1*-flanking P1 Contigs.

Various approaches were employed to extend from the flanking P1 contigs:

YAC AFLP fragments specific to the South end of YAC CIC12F04 (unique to CIC12F04, not present in CIC 12H07) were used to identify P1 clones by AFLP screening of pools of the library.

1. YAC AFLP fragments from YAC 10G07 and overlapping with P1-4 were used to identify P1 clones by AFLP screening of pools of the P1 library.
2. EcoRI restriction fragments from P1 clone P1-6 (resulting from the AFLP-based P1 library screening of step 1) were used as hybridization probes on filters of the BAC library.

Various P1 and BAC clones resulted from this screening and all were AFLP-fingerprinted with the ECs EcoRI/MseI and HindIII/MseI using primers without selective nucleotides. A new map was constructed as described above and is depicted in Figure 7. Table 8 shows the various AFLP PCs having AFLP fragments mapped to flanking YACs and used to screen the P1-library for corresponding P1 clones.

Table 8 represents the various AFLP PCs used to screen the P1 library. The top half of the table shows PCs specific for the North YACs and the bottom half shows the PCs specific for the South YACs. Also indicated are the YACs and P1 clones wherein the AFLP fragments were detected.

Table 8

AFLP PCs	CIC YACs	P1-clones	Comments
PstI-AA MseI-TT	12F04 and 12H07	P1-1, P1-2	Marker Ws84.2
PstI-AT MseI-GT	12F04-specific	P1-1, P1-2	
PstI-CA MseI-AA	12F04-specific	P1-6	
PstI-AC MseI-TG	12F04-specific	P1-7	
PstI-AG MseI-TG	12F04-specific	P1-7	
PstI-CT MseI-GT	12F04-specific	P1-7	
EcoRI-CT MseI-GTA	10G07 and 7E03	P1-3, P1-4	Marker Ws85.1
EcoRI-GT MseI-CTT	10G07 and 7E03	P1-4	Marker Ws86.1
EcoRI-AA MseI-GT	10G07 and 7E03	P1-4, P1-9	
EcoRI-AT MseI-GA	10G07 and 7E03	P1-4, P1-9	
EcoRI-GG MseI-CT	10G07 and 7E03	P1-4, P1-9	

A P1/BAC contig of about 250 kb was obtained covering the South end of YAC CIC12F04 (not extending from this YAC) and containing marker W84.2. A P1 contig of about 150 kb containing markers W85.1 and W86.1 was obtained; this contig is completely contained within YAC CIC7E03.

Construction of a P1/BAC contig covering the *NIM1* gene AFLP marker analysis on the recombinants with markers from the South end of the North P1/BAC contig (WL84.4 and WL84.5, see below and table 11) showed that the previous "walking" steps were unsuccessful in the construction of a contig containing the *NIM1* gene (see next section). Therefore, the existing North P1/BAC contig was extended South with the purpose of "walking" across the *NIM1* gene, which would enable the definition and isolation of a specific DNA segment containing the *NIM1* gene. A hybridization-based approach was followed in which P1 or BAC clones located at the South end of the North P1/BAC contig

were used to identify clones positioned closer to *NIM1* (South bound). New clones resulting from the walking steps were mapped with respect to the existing contigs using AFLP fingerprinting with the ECs EcoRI/MseI and HindIII/MseI as described above. A total of 5 subsequent walking steps appeared to be necessary to "cross" the *NIM1* gene. Table 9 shows the clones obtained in the various walking steps.

Table 9 is an overview of the various walking steps showing the hybridization probe used to screen the P1 and BAC libraries and the selected clones hybridizing to the probes and extending in the South direction.

Table 9

	Probe	New clones extending South
Step 1:	P1-7	BAC-02
Step 2:	BAC-02	P1-16, BAC-03
Step 3:	BAC-03	P1-17, P1-18
Step 4:	P1-18	P1-21, P1-20, BAC-04
Step 5:	BAC-04	P1-22, P1-23, P1-24, BAC-06, BAC-05

A physical map of the various clones resulting from this walking effort is depicted in Figure 8. A total distance of about 600 kb was covered starting from the initial walking point marker W84.2. The South end of the contig presented in Figure 8 appeared to contain the *NIM1* gene (see next section). The contig extends more than 300 kb South from YAC CIC12F04 and appeared not to overlap with YACs CIC10G07 and CIC7E03, indicating that the *NIM1* gene is in the gap between the flanking YAC contigs and that this gap is at least 300 kb.

vi. Construction of an Integrated Genetic and Physical Map of the *NIM1* Region.

In the previous sections it was described how AFLP markers linked to the *NIM1* region were isolated, how YACs corresponding to the flanking markers were identified, and how a P1/BAC contig was constructed extending about 550 kb South from the closest North flanking AFLP marker W84.2. This section describes the generation of new AFLP markers from the P1/BAC contig, the physical mapping of these markers on this contig and the genetic mapping of these markers with the available recombinants.

1. Generation of New AFLP Markers from the P1/BAC Contig

As described in the previous section, the P1 and BAC clones of the contig extension were characterized by AFLP fingerprinting using the ECs EcoRI/MseI and HindIII/MseI. This defined quite accurately the extent of the overlaps between the various P1 and BAC clones and, in addition, generated a number of AFLP fragments specific for these clones. AFLP primers without selective nucleotides are used in fingerprinting of purified plasmid DNA of the P1 or BAC clones. Selective nucleotides will be necessary, however, to be able to use these P1 or BAC-specific AFLP fragments for detection in *Arabidopsis*. By determining the end sequences of the amplified restriction fragments, AFLP primers having the appropriate selective bases can be designed to amplify the P1- or BAC-specific AFLP fragment in *Arabidopsis*. All AFLP fragments originate from the ecotype Columbia (Col) and, therefore, it also should be determined if the Columbia AFLP markers are informative in the *NIM1* recombinants which are derived from a cross of the ecotypes Landsberg erecta (Ler) and a *nim1* mutant of the ecotype Wassilewskija (Ws-*nim*). In principle, there are 4 types of AFLP fragments, two of which are useful markers as indicated in Table 10 below:

Table 10 is an overview of the types of AFLP markers found. (+) or (-) indicates presence or absence of the AFLP fragment.

Table 10

Col	Ler	Ws-*nim*	marker-type
+	+	+	not informative
+	+	-	Ler marker
+	-	+	Ws marker
+	-	-	not informative

In general, fingerprinting of the P1 and BAC clones generated 30 to 40 EcoRI/MseI AFLP fragments and 60 to 80 HindIII/MseI AFLP fragments for each individual clone. The end sequences of individual fragments were determined by standard sequencing techniques. Next, specific AFLP primers sets with selective extensions of 3 nucleotides for both the EcoRI or HindIII primer and the MseI primer were tested on the following panel of DNAs:

1. P1/BAC clone from which the AFLP marker was derived

2a. Yeast

2b. YAC clone CIC12F04 (only for AFLP fragments from P1-7) 2c. YAC clone CIC10G07

3a. Col, origin of the P1 and BAC libraries

3b. Ler, parent 1 of the *nim* recombinants

3c. Ws-nim, parent 2 of the *nim* recombinants

Six clones were selected for sequence analyses of their EcoRI/MseI and HindIII/MseI AFLP fragments: BAC-01/P1-7, P1-17/P1-18, BAC-04/BAC-06. The AFLP fragments from clone P1-7 were all detected in YAC CIC12F04, indicating that this clone is completely contained within this YAC. None of the P1/BAC-specific AFLP fragments was detected in YAC clone CIC10G07, indicating that the P1/BAC contig does not bridge the gap between the two flanking YAC contigs. AFLP markers selected for analysis of the *nim* recombinants are depicted in Table 11.

Table 11 is an overview of the selected AFLP markers from the AFLP PCs specific for the various P1 and BAC clones. A "WL" marker is a marker originating from the same PC and displaying two AFLP markers, a Ws and a Ler marker, which appeared to be completely linked in repulsion phase upon analysis of the *NIM* recombinants.

Table 11

Origin	Marker name	AFLP primers combination
P1-7	WL84.4	EcoRI-AGC MseI-ACT
P1-7	WL84.5	HindIII-CTC MseI-TTC
P1-17/P1-18	Ler84.6a	HindIII-CGT MseI-ATT
P1-17/P1-18	Ler84.6b	HindIII-ATT MseI-CAT
P1-18	Ler84.6c	HindIII-TCT MseI-TAT
P1-18	Ler84.7	EcoRI-AAA MseI-AGA
BAC-04/06	Ler84.8	EcoRI-TTC MseI-AGT
BAC-06	Ler84.9a	EcoRI-AAA MseI-TGT
BAC-06	Ler84.9b	EcoRI-ATC MseI-TCC
BAC-06	Ler84.9c	EcoRI-ATG MseI-GTA

2. Physical Mapping of the New AFLP Markers.

The AFLP markers described above were physically mapped by detecting their presence in the various P1 and BAC clones. The results are presented in Figures 9-11.

3. Genetic Mapping of the New AFLP Markers.

The AFLP markers were all analyzed on a selected set of the recombinants. The results obtained are summarized in Tables 12a, 12b and 12c.

Table 12a

NIM RECOMBINANTS NORTH OF WL84.4&5

Nr.	Plant	PR-1 on/off	Ler 84	Ws 84.2	WL 84.4&5	Ler 84.6a	Ler 84.6b	Ler 84.6c	Ler 84.7	nim	Ler 85	Ler 84.8	Ler 84.9b	Ler 84.9a	Ler 84.9c	Ws 85.1	Ws 86.1	Ler 86
N1	A-74 nim	off	(H)	H	W			W	W	W	W	W			W	W	W	W
N2	A-113		H	H	W			W	W	W	W	W			W	W	W	W
N3	B-023 Rnim	off	H	H	W			W	W	W	W	W			W	W	W	W
N4	B-297 notnim	on	W	W	H	(H)	H	H	H	H	H	H	H	H	H	H	H	H
N5	B-292 nim	off	H	H	W			W	W	W	W	W			W	W	W	W
N6	D-269	off	H	H	W			W	W	W	W	W			W	W	W	W

| | | |

N15	H-047 Rnim	off	
		H	
		H	
		W	
		W	
		W	
		W	
		W	
		W	
		W	
		W	
		W	
		W	
		W	
		W	
		W	
N16	H-097 (nim)	off	
		H	
		H	
		W	
		W	
		W	
		(W)	
		W	
		W	
		W	
		W	

Tabl 12b

NIM RECOMBINANTS SOUTH OF Ler84.9c

Nr.	Plant	PR-1 on/ off	Ler 84	Ws 84.2	WL 84.4&5	Ler 84.6a	Ler 84.6b	Ler 84.6c	Ler 84.7	nim	Ler 85	Ler 84.8	Ler 84.9b	Ler 84.9a	Ler 84.9c	Ws 85.1	Ws 86.1	Ler 86
S6	A-116 Rnim	off	W	W	W			W	W	W	W	W	W	W	W	H	H	H
S7	B-111 nim	off	W	W	W			W	W	W	W	W	W	W	W	H	H	H
S8	B-165 notnim	interm	H	H	H	H	H	H	H	H	H	H	H	H	H	W	W	W
S9	B-182 notnim	interm.	H	H	H	H	H	H	H	H	H	H	H	H	H	W	W	W
S10	B-190 segWT	off	W	W	W			W	W	(W)	W	W	W	W	W	H	H	H

SUBSTITUTE SHEET (RULE 26)

S11	B-243 nim	off	W	W	W			W	W	W	W	W	W	W	H	H	H	H	H
S12	C-036 nim	off	W	W	W			W	W	W	W	W	W	W	H	H	H	H	H
S13	C-088 Rnim	off	W	W	W			W	W	W	W	W	W	W	H	H	H	H	H
S14	D-249 nim	off	W	W	W			W	W	W	W	W	W	W	H	H	H	H	H
S15	E-050 nim	off	W	W	W	W	W	W	W	W	W	W	W	W	H	H	H	H	H
S16	G-058 nim		W	W	W	W	W	W	W	W	W	W	W	W	H	H	H	H	H
S17	C-017 Rnim?		W	W				W	W	W	W	W	W	W	W	W	H	(H)	(H)
S18	H-083		W	W				W	W		W	W	W	W	W	W	H	(H)	(H)

Table 12c

NIM RECOMBINANTS BETWEEN WL84.4&5 AND Ler84.9c

Nr.	Plant	PR-1 on/ off	Ler 84	Ws 84.2	WL 84.4&5	Ler 84.6a	Ler 84.6b	Ler 84.6c	Ler 84.7	nim	Ler 85	Ler 84.8	Ler 84.9b	Ler 84.9a	Ler 84.9c	Ws 85.1	Ws 86.1	Ler 86
N17	B-304 nim		H	H	H	W	W	W	W	W	W	W	W	W	W	W	W	W
N18	C-111 nim	off	H	H	H	W	W	W	W	W	W	W	W	W	W	W	W	W
N19	E-093 nim	off	H	H	H	W	W	W	W	W	W	W	W	W	W	W	W	W
N20	E-110 Rnim	off	H	H	H	W	W	W	W	W	W	W	W	W	W	W	W	W
N21	G-014 nim	off	H	H	H	W	W	H	W	W	W	W	W	W	W	W	W	W
N22	A-019	int	W	W	W	W	H	H	H	H	H	H	H	H	H	--!	H	H

SUBSTITUTE SHEET (RULE 26)

	notnim	erm																		
N23	C-074 Rnim	off	H	H	H	H	H	W	W	W	W	W	W	W	W	W	W	W	W	W
N24	D-169 nim	off	H	H	H	H	H	W	W	W	W	W	W	W	W	W	W	W	W	W
N25	E-103 nim	off	H	H	H	H	H	W	W	W	W	W	W	W	W	W	W	W	W	W
S1	C-105 nim	off	W	W	W	W	W	W	W	W	W	W	H	H	H	H	H	H	H	H
S2	H-039 nim		W	W	W	W	W	W	W	W	W	W	W	H	H	H	H	H	H	H
S3	B-052 nim	off	W	W	W	W	W	W	W	W	W	W	W	W	H	H	H	H	H	H
S4	B-142 nim	off	W	W	W	W	W	W	W	W	W	W	W	W	W	H	H	H	H	H
S5	B-110 Rnim	off	W	W	W	W	W	W	W	W	W	W	W	W	W	W	H	H	H	H

The AFLP markers Ler84.8, Ler84.9a, Ler84.9b and Ler84.9c appeared to map at the South side of *NIM1*. Recombinants were found that were phenotypically *nim1* (homozygous, genotype Ws-*nim1*/Ws-*nim1*) and heterozygous for these AFLP markers (the Ler-specific AFLP marker was detected, genotype is Ws-*nim1*/Ler). AFLP marker Ler84.8 appeared to be closest to *NIM1*: only a single recombinant (C-105) was scored as heterozygous Ws-*nim1*/Ler and homozygous Ws-*nim1*/Ws-*nim1*. AFLP markers Ler84.7 and Ler84.6c appeared to completely cosegregate with *NIM1*: all recombinants had an identical *NIM1* and AFLP marker genotype. North of *NIM1*, marker L84.6b appeared to be closest to *NIM1*: three *nim1* phenotype recombinant plants, C-074, D-169 and E-103 (Table 12c), were found to be heterozygous Ws-*nim1*/Ler at this marker. With the aid of the cosmid contig generated from P1-18, BAC-04 and BAC-06, AFLP markers Ler84.6b and Ler84.8 were mapped in P1-18 and BAC-04, respectively, and found to have a physical distance of approximately 110 kb. This defines *nim1* to be located on a DNA segment estimated to be 110 kb in length. From this analysis it has been determined that the *NIM1* gene is contained in clone BAC-04 or P1-18. Clones BAC-04 and P1-18 have been deposited with ATCC and given deposit numbers ATCC 97543 and ATCC 97606, respectively.

vii. Genetic and Physical Fine Mapping of the *NIM1* Gene.

The previous section described how a DNA segment containing the *NIM* gene was delineated by physical mapping of the flanking AFLP markers (Ler84.6b and Ler84.8) on the P1/BAC contig. The flanking markers appeared to map on two overlapping clones, P1-18 and BAC-04. This section describes how additional BAC-04-specific and P1-18-specific AFLP markers were generated to increase the resolution of the genetic and physical map in the region containing the *NIM1* gene.

viii. Generation of Additional AFLP Markers from the Cosmid Array.

Four ECs were selected to generate additional AFLP markers for fine mapping of *NIM1*: PstI/MseI, XbaI/MseI, BstYI/MseI and TaqI/MseI. PstI/MseI and XbaI/MseI AFLP fragments were generated on clone P1-18 and BAC-04 and the selective sequences necessary for detection in Arabidopsis were determined. Similarly, the AFLP fragments and selective sequences were determined for BstYI/MseI and TaqI/MseI; however, in this case the procedure was performed using cosmid DNAs: A11, C7, E1 and E8 for BstYI/MseI (complete *NIM1* region) and D7, E8 and E6 for TaqI/MseI (South side of *NIM1* region). Informative AFLP markers selected for further genetic and physical mapping are shown in Table 13. Additional adapters used in this work are shown in Table 14.

Table 13 shows the AFLP markers used for genetic and physical fine mapping of *NIM1*. "BstYI(T)" indicates that the restriction site and corresponding primer was either AGATCT or GGATCT.

Table 13

Marker	EC/PC	
Ler84.Y1	BstYI(T)-GCT	MseI-AAC
Ws84.Y2	BstYI(T)-TCT	MseI-GCA
Ler84.Y3	BstYI(T)-AAG	MseI-TAT
Ler84.Y4	BstYI(T)-GTT	MseI-AGA
Ws84.T1	TaqI-TAC	MseI-GGA
Ler84.T2	TaqI-TTG	MseI-GGA

Table 14 shows same Additional adapters used for identifying new AFLP markers.

Table 14

BstYI:	5'-CTCGTAGACTGCGTACC-3'
	3'-CATCTGACGCATGGCTAG-5'
TaqI:	5'-CTCGTAGACTGCGTACC-3'
	3'-CATCTGACGCATGGGC-5'

ix. Physical mapping of new AFLP markers to the cosmid contig.

The markers indicated above were physically mapped on the cosmid array by determining their presence in the various cosmid clones (Figure 11).

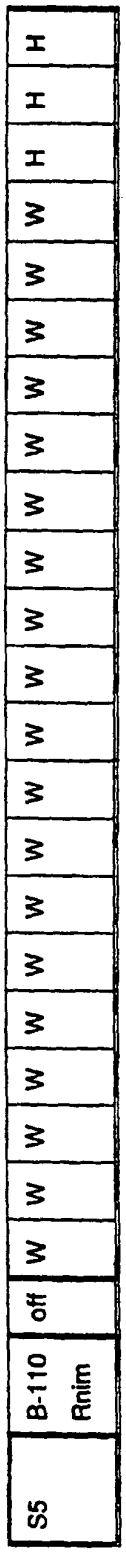
1. Genetic Mapping of New AFLP Markers.

The new AFLP markers were genetically mapped by AFLP analysis of the closest North and South recombinants. The closest North (recombinant D169) and South (recombinant C105) recombination points were mapped (see Table 15). The AFLP analysis showed that recombinant D169 had a recombination South of marker L84.Y1, but North of marker W84.Y2. The recombination point in recombinant C105 mapped between markers L84.T2 and L84.8. Using the available set of recombinants this allowed further delineation of the chromosomal interval containing *NIM1*; the distance between the flanking recombination points appeared to be 60-90 kb (Figure 12).

Tabl 15 NIM RECOMBINANTS BETWEEN WL84.4&5 AND Ler84.9c

Nr.	Plant	P R-1 on / off	Le r 84	W s 84 .2	W L 84 . 4& 5	Le r8 4. 6a	Le r8 4. 6b	Le r8 4. Y1	W s 84 .Y 2	Le r 84 .7	Le r 84 6c	Le r8 4. Y3	W s 84 .Y 4	Le r8 4. T1	Le r8 4. T2	Le r 85	Ni m	Le r8 4. 8	Le r8 4. 9b	Le r8 4. 9a	Le r8 4. 9c	W s 85 .1	W s 86 .1	Le r 86
N17	B-304 nim		H	H	H	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
N18	C-111 nim	off	H	H	H	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
N19	E-093 nim	off	H	H	H	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
N20	E-110 Rnim	off	H	H	H	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
N21	G-014 nim	off	H	H	H	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W

| | | |


S5
B-110
Rnim
off
W
W
W
W
W
W
W
W
W
W
W
W
W
W
W
W
W
W
W
H
H
H

2. Construction of a Cosmid Contig.

For complementation of the *nim1* plant phenotype, transformation of *nim1* plants is required with a wildtype *NIM1* gene. This can be accomplished by transforming these plants with a cosmid containing the gene. For this purpose, a cosmid contig of the *NIM1* region is constructed. Since *Arabidopsis* is transformed using *Agrobacterium*, the cosmid vector used is a binary vector.

DNA was isolated from BAC-04, BAC-06 and P1-18, and used to make a partial digest using restriction enzyme Sau3AI. The 20-25 kb fragments were isolated using a sucrose gradient, pooled, and filled in with dATP and dGTP. The binary vector (04541) was cleaved with XhoI and filled in with dCTP and dTTP. The fragments were next ligated into the vector. The ligation mix was packaged and transduced into *E.coli*.

This cosmid library was screened with the BAC-04, BAC-06 and P1-18 clones and positive clones isolated. These cosmids were next AFLP fingerprinted and arranged into a contig of overlapping clones spanning the *NIM1* region. The insert sizes of the cosmids were determined, and limited restriction mapping was performed. The results are shown in Figure 10.

Example 2

Identification of a Clone Containing the *NIM1* gene.

1. Complementation Via Stable Transformation

Cosmids that are generated from clones that span the *NIM1* region (described above) are moved into *Agrobacterium* by triparental mating. These cosmids are then used to transform *nim1 Arabidopsis* by vacuum infiltration (Mindrinos et al., 1994, Cell 78, 1089-1099) or by standard root transformation. Seed from these plants is harvested and allowed to germinate on agar plates with kanamycin (or another appropriate antibiotic) as selection agent. Only plantlets that are transformed with cosmid DNA can detoxify the selection agent and survive. Seedlings that survive the selection are transferred to soil and tested for the *nim* phenotype or their progeny are tested for the *nim* phenotype. Transformed plants that no longer have the *nim* phenotype identify cosmid(s) that contain a functional *NIM1* gene.

2. Complementation in a Transient Expression System.

The ability of DNA clones to complement the *nim1* mutation is tested in 2 transient expression systems.

In the first system, *nim1 Arabidopsis* plants containing a PR1-luciferase (PR1-lux) transgene are used as bombardment recipient material. These plants are generated by transforming Columbia ecotype plants with a PR1-lux construct by vacuum infiltration, followed by kanamycin selection of the harvested seed, as described above. Transformed plants that express luciferase activity after induction with INA are selfed and homozygous plants are generated. These are crossed to *nim1* plants. In the transient assay, progeny plants from this cross that are homozygous for *nim1* and for PR1-lux are used for identification of DNA clones that can complement the *nim1* phenotype. To this end, the plants are first treated with INA, as described in example 1.1 above. Two days later these

plants are harvested, surface sterilized and plated on GM agar medium. The leaf tissue is then bombarded with cosmid, P1 or BAC clones (or subclones) from the *NIM1* region and after one day, the luciferase activity of the leaves is measured. Clones that induce luciferase activity contain the *NIM1* gene.

In a second system, *nim1* plants are treated with INA (as described in example 1.1 above) and 2 days later bombarded with cloned DNA (cosmid, P1, BAC and/or YAC clones or subclones) from the *NIM1* locus region and a reporter plasmid. The reporter plasmid contains the luciferase gene, driven by the *Arabidopsis* PR1 promoter (PR1-lux). In *nim1* plants, INA does not activate the PR1 promoter (as described in example 1.2 above) and thus can not induce luciferase activity from the reporter plasmid. However, when a cotransformed DNA clone contains the complementing *NIM1* gene, INA does induce the PR1 promoter, as evidenced by an induction of luciferase activity. One day after the cobombardment, the luciferase activity of the whole plant is measured. DNA clones (cosmids, P1 or BAC clones or subclones) that induce luciferase activity that is significantly above background levels contain the *NIM1* gene.

3. Changes in transcripts in *nim1* phenotype lines.

Since *nim1* phenotype plants have mutations in the *NIM1* gene, it is conceivable that in some lines the gene is altered in such a manner that there is no mRNA transcribed, or an aberrant mRNA (size) is produced. To test for this, RNA blot analysis is performed on the *nim1* lines.

RNA is isolated from Ws and Ler plants of these lines, (after water or INA or BTH treatment) and used to prepare northern blots. These blots are hybridized with DNA fragments isolated from clones of the DNA contig of the *NIM1* locus. DNA fragments that identify *nim1* lines with aberrant RNA expression (aberrant in size or concentration), likely

identify (part of) the *NIM1* gene. The DNA fragment and surrounding DNA is sequenced and used to isolate a cDNA (by library screen or by reverse transcription-PCR), which is also sequenced. The clone from which the fragment was isolated or the isolated cDNA is used to show complementation of the *nim1* phenotype in stable and transient expression systems.

Example 3

Determination of the DNA sequence of the *NIM1* gene.

1. Genomic sequencing.

Genomic clones that may contain the *NIM1* gene are sequenced using methods known in the art. These include BAC-04, P1-18 and the cosmids from the *NIM1* region. For instance, the cosmids are digested with restriction enzymes and fragments that are derived from the insert are cloned into a general purpose vector, such as pUC18 or Bluescript. The larger P1 and BAC clones are randomly sheared and fragments cloned into a general purpose vector. The fragments in these vectors are sequenced by conventional methods (e.g. by "primer walking" or generation of deletions of inserts). The obtained sequences are assembled into a contiguous sequence.

The sequence of the insert of a complementing clone contains the *NIM1* gene. The approximate start and end of the *NIM1* gene are deduced based on the DNA sequence, sequence motifs such as TATA boxes, the open reading frames present in the sequence, codon usage, the cosmid complementation data, the relative location of the AFLP markers and additional relevant data that is gathered (see Example 4, below).

2. cDNA sequencing.

The cosmid(s) or larger clones that contain the *NIM1* gene (as described in Example 2 above), are used to isolate cDNAs. This is accomplished by using the clones (or DNA fragments) as probes in a screen of a cDNA library of wildtype *Arabidopsis* plants. The cDNAs that are isolated are sequenced as described for cosmid sequencing and used in complementation tests. To this end, full length cDNAs are cloned into a suitable plant expression vector, behind a constitutive promoter. These constructs are used in the transient assays as described above. Alternatively, the cDNAs are cloned into a binary expression vector, allowing for expression in plant tissues and for *Agrobacterium*- mediated plant transformation, as described in Example 2 above. A cDNA that contains the *NIM1* gene (as determined by complementation, isolation with a closely linked AFLP marker, isolation with a cosmid fragment, or by other deduction) is sequenced.

The genes from Ws-O and *nim1* plants are isolated and sequenced. The genes are obtained from a cosmid of cDNA library, using a fragment of the isolated *NIM1* gene as a probe. Alternatively, the genes or cDNAs are isolated by PCR, using *NIM1*-gene-specific primers and genomic DNA or cDNA as template. Likewise, the *nim1* alleles from other *nim1* lines (see Example 1.1 above) are isolated and sequenced in a similar manner.

Example 4

Description of the *NIM1* gene and deduced protein sequence

The DNA sequence of the *NIM1* gene or cDNA is determined as described in Example 3 above. This sequence is analyzed with the use of DNA analysis programs, such as can be found in the Genetics Computer Group (GCG) package, in the Sequencer or Staden packages, or any similar DNA analysis program package.

Specifically, the start and end of the gene are determined, based on open reading frame analysis, the presence of stop and potential start codons, the presence of potential promoter motifs (such as the TATA box), the presence of polyadenylation signals and the like. Also, the predicted amino acid is deduced from the open reading frame. Both the DNA and protein sequence are used to search databases for sequences with homologies, such as transcription factors, enzymes or motifs of such genes or proteins.

Example 5

Isolation of *NIM1* homologs

The *Arabidopsis NIM1* gene may be used as a probe in the low stringency hybridization screening of a genomic or cDNA library in order to isolate *NIM1* homologs from other plant species. Alternatively, this is accomplished by PCR amplification, using primers designed based on the *Arabidopsis NIM1* gene sequence and using genomic DNA or cDNA as template. The *NIM1* gene may be isolated from corn, wheat, rice, barley,rape seed, sugarbeet, potato, tomato, bean, cucumber, grape, tobacco and other crops of interest and sequenced. With a set of sequences from *NIM1* gene homologs in hand, new primers can be designed from conserved portions of the gene, in order to isolate *NIM1* homologs from more distantly related plant species by PCR amplification.

Example 6

Complementation of the *nim1-1* gene with genomic fragments.

1. Construction of a cosmid contig.

A cosmid contig of the *NIM1* region was constructed using CsCl-purified DNA from BAC04, BAC06 and P1-18. The DNAs of the three clones were mixed in equimolar quantities and were partially digested with the restriction enzyme Sau3A. The 20-25 kb

fragments were isolated using a sucrose gradient, pooled and filled in with dATP and dGTP. Plasmid pCLD04541 was used as T-DNA cosmid vector. This plasmid contains a broad host range pRK290-based replicon, a tetracycline resistance gene for bacterial selection and the nptII gene for plant selection. The vector was cleaved with XhoI and filled in with dCTP and dTTP. The prepared fragments were then ligated into the vector . The ligation mix was packaged and transduced into *E. coli* strain XL1-blue MR (Stratagene). Resulting transformants were screened by hybridization with the BAC04, BAC06 and P1-18 clones and positive clones isolated. Cosmid DNA was isolated from these clones and template DNA was prepared using the ECs EcoRI/MseI and HindIII/MseI. The resulting AFLP fingerprint patterns were analyzed to determine the order of the cosmid clones. A set of 15 semi-overlapping cosmids was selected spanning the *nim* region (Figure 13). The cosmid DNAs were also restricted with EcoRI, PstI, BssHII and SgrAI. This allowed for the estimation of the cosmid insert sizes and the verification of the overlaps between the various cosmids as determined by AFLP fingerprinting.

2. Identification of a clone containing the *NIM* gene.

Cosmids generated from clones spanning the *NIM1* region were moved into *Agrobacterium tumefaciens* AGL-1 through conjugative transfer in a tri-parental mating with helper strain HB101 (pRK2013). These cosmids were then used to transform a kanamycin-sensitive *nim1-1* Arabidopsis line using vacuum infiltration (Mindrinos et al., 1994, Cell 78, 1089-1099). Seed from the infiltrated plants was harvested and allowed to germinate on GM agar plates containing 50 mg/ml kanamycin as a selection agent. Only plantlets that are transformed with cosmid DNA can detoxify the selection agent and survive. Seedlings that survive the selection were transferred to soil approximately two weeks after plating and

tested for the *nim1* phenotype as described below. Transformed plants that no longer have the *nim1* phenotype identify cosmid(s) that contain a functional *NIM1* gene.

3. Testing for the *nim1* phenotype of transformants.

Plants transferred to soil were grown in a phytotron for approximately one week after transfer. 300µm INA was applied as a fine mist to completely cover the plants using a chromister. After two days, leaves were harvested for RNA extraction and PR-1 expression analysis. The plants were then sprayed with *Peronospora parasitica* (isolate EmWa) and grown under high humidity conditions in a growing chamber with 19°C day/17° night temperatures and 8h light/16h dark cycles. Eight to ten days following fungal infection, plants were evaluated and scored positive or negative for fungal growth. Ws and *nim1* plants were treated in the same way to serve as controls for each experiment.

Total RNA was extracted from the collected tissue using a LiCl/phenol extraction buffer (Verwoerd, et al. NAR 17:2362). RNA samples were run on a formaldehyde agarose gel and blotted to GeneScreen Plus (DuPont) membranes. Blots were hybridized with a ^{32}P-labeled PR-1 cDNA probe. The resulting blots were exposed to film to determine which transformants were able to induce PR-1 expression after INA treatment. The results are summarized in Table 16.

Table 16 shows complementation of the *nim1* phenotype by cosmid clones.

Table 16

Clone Name	# of transformants	# of plants with INA induced PR-1/total # of plants tested (%)
A8	3	0/3 (0%)
A11	8	4/18 (22%)
C2	10	1/10 (10%)
C7	33	1/32 (3%)
D2	81	4/49 (8%)
D5	6	5/6 (83%)
E1	10	10/10 (100%)
D7	129	36/36 (100%)
E8	9	0/9 (0%)
F12	6	0/6 (0%)
E6	1	0/1 (0%)
E7	34	0/4 (0%)
WS-control (wild-type)	NA	28/28 (100%)
nim1-1 phenotype control	NA	0/34 (0%)

NA-not applicable

Example 7

Sequencing of the 9.9 Kb *NIM1* gene region.

BAC04 DNA (25 ug, obtained from KeyGene) was the source of DNA used for sequence analysis. This BAC was shown to be the clone completely encompassing the region that complemented the *nim1* mutants. DNA was randomly sheared using an

approach from Cold Spring Harbor. Briefly, BAC DNA was sheared in a nebulizer to an average molecular weight of about 2 kb. Ends of the sheared fragments were repaired using a two-step protocol with dNTPS, T4 DNA polymerase and Klenow fragment (Boehringer). The end-repaired DNA was run on a 1% low-melt agarose gel and the region between 1.3 kb and 2.0 kb was cut from the gel. DNA was isolated from the gel fragment by a freeze-thaw approach. DNA was then mixed with EcoRV-digested pBRKanF4 and was ligated overnight at 4°C. pBRKanF4 is a derivative of pBRKanF1, which was obtained from Kolavi Bhat at Vanderbilt University (Bhat, K.S., Gene 134(1), 83-87 (1993)). *E. coli* strain DH5a was transformed with the ligation mix, and the transformation mix was plated onto plates containing kanamycin and X-gal. 1600 white or light blue KanR colonies were selected for plasmid isolation. Individual colonies were picked into 96-well deep well plates (Polyfiltronics, #U508) containing 1.5 ml of TB + Kan (50 ug/ml). Plates were covered and were placed on a rotating platform shaker at 37°C for 16 hrs. Plasmid DNA was isolated using the Wizard Plus 9600 Miniprep system (Promega, #A7000) according to manufacturer's recommendations.

Plasmids were sequenced using Dye Terminator chemistry (Applied BioSystems PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit, P/N 402078) and primers designed to sequence both strands of the plasmids. Data was collected on ABI 377 DNA sequencers. Approximately 75% of these reactions yielded useful sequence information. Sequences were edited and were assembled into contigs using Sequencher 3.0 (Gene Codes Corporation), Staden gap4 (Roger Staden, e-mail address rs@mrc-lmb.cam.ac.uk), and PHRED (Phil Green, e-mail address phg@u.washington.edu). The largest contig (approximately 76 kb) covered the complementing region to an average depth of 7 independent calls/base.

A region of approximately 9.9 kb defined by the overlap of cosmids E1 and D7 was identified by complementation analysis to contain the *nim1* region. Primers which flanked the insertion site of the vector and specific to the cosmid backbone were designed using Oligo 5.0 Primer Analysis Software (National Biosciences, Inc.). DNA was isolated from cosmids D7 and E1 using a modification of the ammonium acetate method (Traynor, P.L., 1990. BioTechniques 9(6): 676.) This DNA was directly sequenced using Dye Terminator chemistry above. The sequence obtained allowed determination of the endpoints of the complementing region.

A truncated version of the BamHI-EcoRV fragment was also constructed, resulting in a construct which contains none of the "Gene 3" region (Fig. 13). The following approach was necessary due the presence of HindIII sites in the Bam-Spe region of the DNA. The BamHI-EcoRV construct was completely digested with SpeI, then was split into two separate reactions for double digestion. One aliquot was digested with BamHI, the other HindIII. A BamHI-SpeI fragment of 2816 bp and a HindIII-SpeI fragment of 1588 bp were isolated from agarose gels (QiaQuick Gel extraction kit) and were ligated to BamHI-HindIII-digested pSGCG01. DH5a was transformed with the ligation mix. Resulting colonies were screened for the correct insert by digestion with HindIII following preparation of DNA using Wizard Magic MiniPreps (Promega). A clone containing the correct construct was electroporated into *Agrobacterium* strain GV3101 for transformation of *Arabidopsis* plants.

Example 8

Identification of the *NIM1* gene region by allele sequencing.

Table 17. Genetic segregation of non-inducible immunity mutants.

				Phenotype	
Mutant	Generation	Female	Male	Wildtype	*nim1*
nim1-1[a]	F1	wildtype[b]	*nim1-1*	24	0
	F2			98	32
nim1-2	F1	*nim1-2*	Wildtype	3	0
nim1-3	F1	*nim1-3*	Wildtype	3	0
nim1-4	F1	*nim1-4*	Wildtype	3	0
nim1-5	F1	*nim1-5*	Wildtype	0	35
nim1-6	F1	*nim1-6*	Wildtype	3	0
nim1-2	F1	*nim1-2*	*nim1-1*	0	15
nim1-3	F1	*nim1-3*	*nim1-1*	0	10
nim1-4	F1	*nim1-4*	*nim1-1*	0	15

nim1-5	F1	*nim1-5*	*nim1-1*	0	14
	F2			9	85
nim1-6	F1	*nim1-6*	*nim1-1*	0	12

[a] Data from Delaney et al. (1995) PNAS 92,6602-6606.

[b] Wild type denotes the wildtype Ws-0 strain.

---

1. Genetic Analyses

To determine dominance of the various mutants that displayed the *nim1* phenotype, pollen from wildtype plants was transferred to the stigmata of *nim1-1, -2, -3, -4, -5, -6.* If the mutation is dominant, then the *nim1* phenotype will be observed in the resulting F1 plants. If the mutation is recessive, then the resulting F1 plants will exhibit a wildtype phenotype.

The data presented in Table 17 show that when *nim1-1, -2, -3, -4* and *-6* are crossed with the wildtype, the resulting F1 exhibit the wildtype phenotype. Thus, these mutations are recessive. In contrast, the *nim1-5 X* wildtype F1 progeny all exhibit the *nim1* phenotype, indicating that this is a dominant mutation. Following INA treatment, no *P. parasitica* sporulation was observed on wildtype plants, while the F1 plants supported growth and some sporulation of *P. parasitica.* However, the *nim1* phenotype in these F1 plants was less severe then observed when *nim1-5* was homozygous.

To determine allelism, pollen from the kanamycin-resistant *nim1-1* mutant plants was transferred to the stigmata of *nim1-2,* -3, -4, -5, -6. Seeds resulting from the cross were plated onto Murashige-Skoog B5 plates containing kanamycin at 25 ug/ml to verify the hybrid origin of the seed. Kanamycin resistant (F1) plants were transferred to soil and assayed for the *nim1* phenotype. Because the F1 progeny of the cross of the *nim1-5*

mutant with the Ws wildtype displayed a *nim1* phenotype, analysis of *nim1-5 X nim1-1* F2 was also carried out.

As shown in Table 17, all of the resulting F1 plants exhibited the *nim1-1* phenotype. Thus, the mutation in the *nim1-2, -3, -4, -5, -6* was not complemented by the *nim1-1*; these plants all fall within the same complementation group and are therefore allelic. Analysis of F2 progeny from the *nim1-5 X nim1-1* cross also displayed the *nim1* phenotype, confirming that *nim1-5* is a *nim1* allele.

2. Sequence Analysis and Subcloning of the NIM1 Region

The 9.9 kb region containing the *NIM1* region was analyzed for the presence of open reading frames in all six frames using Sequencher 3.0 and the GCG package. Four regions containing large ORF's were identified as possible genes (Gene regions 1-4). These four regions were PCR amplified from DNA of the wild-type parent and six different *nim1* allelic variants. Primers for these amplifications were selected using Oligo 5.0 (National Biosciences, Inc.) and were synthesized by Integrated DNA Technologies, Inc. PCR products were separated on 1.0% agarose gels and were purified using the QIAquick Gel Extraction Kit. The purified genomic PCR products were directly sequenced using the primers used for the initial amplification and with additional primers designed to sequence across any regions not covered by the initial primers. Average coverage for these gene regions was approximately 3.5 reads/base.

Sequences were edited and were assembled using Sequencher 3.0. Base changes specific to various *nim1* alleles were identified only in the region designated Gene Region 2.

The positions listed in Table 18 relate to Figure 14 and relate to the top strand of the 9.9 kb region featured in Figure 13. The open reading frames from the gene regions described in Figure 13 as 1, 2, 3 and 4 were sequenced and the changes in the different

nim1 alleles are shown in the Table. The changes that are described are on the top strand, 5¢ to 3¢, as it would relate to Figure 13.

It is apparent that the *NIM1* gene was cloned and that it lies within Gene Region 2, since there are amino acid changes or alterations of sequence within the open reading frame of Gene Region 2 in all 6 *nim1* alleles. At the same time, at least one of the *nim1* alleles shows no changes in the open reading frames within Gene Regions 1, 3 and 4. Therefore, the only gene within the 9.9 kb region that could be *NIM1* is the Gene Region 2, the *NIM1* gene.

The Ws section of Table 18 indicates the changes in the Ws ecotype of *Arabidopsis* relative to the Columbia ecotype of *Arabidopsis*. Figures 13, 14, 15 and all others wherein sequence is shown relate to the Columbia ecotype of *Arabidopsis*, which contains the wild type gene in the experiments that were conducted. The changes are listed as amino acid changes within the gene 2 or *NIM1* region and are listed as changes in base pairs in the other regions.

Figure 13 shows 4 different panels that describe the cloning of the *NIM1* gene and describe the entire 9.9 kb region. Figure 14 is the sequence of the entire 9.9 kb region in the same orientation as described in Figure 13. Figure 15 is the sequence of the specific *NIM1* gene region which is gene region 2 indicated in Figure 13; the sequence of Figure 15 contains the *NIM1* gene. Figure 15 shows the amino acid sequence in single letter code and shows the full length cDNA and RACE product that was obtained in capital letters in the DNA sequence. Some of the allele mutations that were found are shown above the DNA sequence and the particular *nim1* allele that had that change is indicated.

Sequence analysis of the region and sequencing of various *nim1* alleles (see below) allowed identification of a cosmid region that contains the *nim1* gene. This region is delineated by a BamH1-EcoRV restriction fragment of ~5.3 kb. Cosmid DNA from D7 and

plasmid DNA from pBlueScriptII(pBSII)were digested with Bam HI and with EcoRV (NEB). The 5.3 kb fragment from D7 was isolated from agarose gels and was purified using the QIAquick gel extraction kit (# 28796, Qiagen). The fragment was ligated overnight to the Bam-EcoRV-digested pBSII and the ligation mixture was transformed into *E. coli* strain DH5a. Colonies containing the insert were selected, DNA was isolated, and confirmation was made by digestion with HindIII. The Bam- EcoRV fragment was then engineered into a binary vector (pSGCG01) for transformation into *Arabidopsis*.

3. Northern analysis of the four gene regions.

Identical Northern blots were made from RNA samples isolated from water-, SA-, BTH- and INA-treated Ws and *nim1* lines as previously described (Delaney et al, 1995, PNAS 92, 6602-6606). These blots were hybridized with PCR products generated from the four gene regions identified in the 9.9 kb *NIM1* gene region. Only the gene region containing the *NIM1* gene (Gene Region 2) had detectable hybridization with the RNA samples, indicating that only the *NIM1* region contains a detectable transcribed gene (Figure I6 and Table 18).

Table 18 shows *nim1* allele sequence variation.

Table 18

	Gene Region			
Allele/ ecotype	1 (bases 590-1090)	2 (*NIM1*) (bases 1380-4100)	3 (bases 5870 - 6840)	4 (bases 8140-9210)
nim1-1	no changes	t inserted at 2981: change of 7AA and premature termination of protein.	no changes	no changes
nim1-2	no changes	g to a at 2799: His to Tyr	no changes	no changes
nim1-3	no changes	deletion of t at 3261: change of 10AA and premature termination of protein.	no changes	no changes
nim1-4	no changes	c to t at 2402: Arg to lys	no changes	no changes
nim1-5	no changes	c to t at 2402: Arg to lys	no changes	no changes
nim1-6	g to a at 734: asp to lys	g to a at 2670: Gln to Stop	no changes	no changes
WS (compared to Columbia)	no changes	a to g at 1607: Ile to Leu a to c at 2344: intron t to g at 2480: Gln to Pro g to c at 2894: Ser to Trp ggc deleted at 3449: lose Ala c to t at 3490: Ala to Thr c to t at 3498: Ser to Asn a to t at 3873: non-coding g to a at 3992: non-coding	t to a at 5746 a to t at 5751 t to a at 5754 c to t at 6728 a to t at 6815 t to c at 6816	t to g at 8705 g to t at 8729 g to t at 8739 g to t at 8784 c to a at 8789 c to t at 8812 a to g at 8829 t to g at 8856 a to c at 9004

		g to a at 4026: non-coding g to a at 4061: non-coding		a to t at 9011 a to g at 8461
RNA detected	No	Yes	No	No

Positions listed in the table relate to Figure 14 containing the 9.9Kb sequence. All alleles *nim1-1* to *nim1-6* are WS strain. Columbia-0 represents the wild type

We also demonstrated that the gene region 2 (Fig. 13) contains the functional *NIM1* gene by doing additional complementation experiments. A BamHI/HindIII genomic DNA fragment containing gene region 2 was isolated from cosmid D7 and was cloned into the binary vector pSGCG01 containing the gene for kanamycin resistance (Fig. 13; Steve Goff, personal communication). The resulting plasmid was transformed into the Agrobacterium strain GV3101 and positive colonies were selected on kanamycin. PCR was used to verify that the selected colony contained the plasmid. Kanamycin-sensitive *nim1-1* plants were infiltrated with this bacteria as prevously described. The resulting seed was harvested and planted on GM agar containing 50µg/ml kanamycin. Plants surviving selection were transferred to soil and tested for complementation. Transformed plants and control Ws and *nim1* plants were sprayed with 300µm INA. Two days later, leaves were harvested for RNA extraction and PR-1 expression analysis. The plants were then sprayed with *Peronospora parasitica* (isolate EmWa) and grown as previously described. Ten days following fungal infection, plants were evaluated and scored positive or negative for fungal growth. All of the 15 transformed plants, as well as the Ws controls, were negative for fungal growth following INA treatment, while the *nim1*controls were positive for fungal growth. RNA was extracted and analyzed as described above for these transformants and controls. Ws controls and all 15 transformants showed PR-1 gene induction following INA treatment, while the *nim1* controls did not show PR-1 induction by INA.

4. Isolation of a *NIM1* cDNA.

An Arabidopsis cDNA library made in the lYES expression vector (Elledge et al, 1991, PNAS 88, 1731-1735) was plated and plaque lifts were performed. Filters were hybridized with a ^{32}P-labeled PCR product generated from the gene region containing nim-1. 14 positives were identified from a screen of approximately 150,000 plaques. Each plaque was purified and plasmid DNA was recovered. cDNA inserts were digested out of the vector using EcoRI, agarose-gel-purified and sequenced. Sequence obtained from the longest cDNA is indicated in Figure 15. To confirm that we had obtained the 5¢ end of the cDNA, a Gibco BRL 5' RACE kit was used following manufacturer's instructions. The resulting RACE products were sequenced and found to include the additional bases indicated in Figure I5. The transcribed region present in both cDNA clones and detected in RACE is shown as capital letters in Figure 15. Changes in the alleles are shown above the DNA strand. Capitals indicate the presence of the sequence in a cDNA clone or detected after RACE PCR.

Example 9

Characterization of the *NIM1* gene

The multiple sequence alignment was constructed using Clustal V (Higgins, Desmond G. and Paul M. Sharp (1989), Fast and sensitive multiple sequence alignments on a microcomputer, CABIOS 5:151-153) as part of the DNA* (1228 South Park Street, Madison Wisconsin, 53715) Lasergene Biocomputing Software package for the Macintosh (1994).

It has been determined that certain regions of the *NIM1* protein are homologous in amino acid sequence to 4 different rice cDNA protein products. The homologies were

identified using the *NIM1* sequences in a GenBank BLAST search. Comparisons of the regions of homology in *NIM1* and the rice cDNA products are shown in Figure 19. The *NIM1* protein fragments show from 36 to 48% identical amino acid sequences with the 4 rice products.

Example I0

Phenotypic characterization of the various *nim1* alleles

1. Analysis of chemical responsiveness in *nim1* alleles.

We analyzed the differences among the various *nim1* alleles in terms of chemical induction of PR gene expression and *Peronospora parasitica* resistance (see Figures I7 and 18).

Mutant plants were treated with chemical inducers and then assayed for PR gene expression and disease resistance.

2. Plant growth and chemical application.

Wild-type seeds and seeds for each of the *nim1* alleles (*nim1-1, -2, -3, -4, -5, -6*) were sown onto MetroMix 300 growing media, covered with a transparent plastic dome and placed at 4° C in the dark for 3 days. After 3 days of 4° C treatment the plants were moved to a phytotron for 2 weeks. At approximately 2 weeks post-planting, germinated seedlings had produced 4 true leaves. Plants were then treated with H_2O, 5mM SA, 300 uM BTH or 300 uM INA. Chemicals were applied as a fine mist to completely cover the seedlings using a chromister. Water control plants were returned to the growing phytotron while the chemically treated plants were held in a separate but identical phytotron. After 3 days plants were divided into 2 groups. One group was harvested for RNA extraction and analysis. The second group was inoculated with *P. parasitica*.

3. *Peronospora parasitica* inoculation and analysis.

P. parasitica isolate 'EmWa' is a *P.p.* isolate that is compatible in the Ws ecotype. Compatible isolates are those that are capable of causing disease on a particular host. The *P. parasitica* isolate 'NoCo' is incompatible on Ws but compatible on the Columbia ecotype. Incompatible pathogens are recognized by the potential host, eliciting a host response that prevents disease development. At 3 days post-chemical application water and chemically treated plants were inoculated with the compatible 'EmWa' isolate. 'NoCo' inoculation was conducted on water treated plants only. Following inoculation plants were covered with a clear plastic dome to maintain high humidity required for successful *P. parasitica* infection and placed in a growing chamber with 19° C day/17° C night temperatures and 8h light/16h dark cycles.

At various timepoints after inoculation plants were analyzed microscopically to assess symptom development. Under magnification sporulation of the fungus can be observed at very early stages of disease development. The percentage of plants/pot showing sporulation at 5d, 6d, 7d, 11d and 14d after inoculation was determined and the density of sporulation was also recorded.

Figure 18 shows the disease assessment of the various *nim1* alleles following *P. parasitica* inoculation. The most distinguishing timepoints are 5 and 6 days post-inoculation. At 5 days post-inoculation *nim1-4* shows ~80% infection under all inducing chemical treatments performed, clearly indicating that this allele/genotype has the most severe disease susceptibility. At 6 days post-inoculation, *nim1, -2, -3, -4* and *-6* show significant disease incidence under all inducing chemical treatments. However, *nim1-5* shows less infection than Ws wild-type under all treatments at day 6. Therefore, *nim1-5* is the most disease resistant of the various *nim1* alleles. *nim1-2* appears intermediate with respect to disease susceptibility after BTH but not the other inducing treatments.

PR-1 gene expression indicates that *nim1-4* is the least responsive to all of the inducing chemicals tested (Figure 17), while *nim1-5* shows elevated levels of PR-1 expression in the absence of inducers. These PR-1 gene expression results are consistent with the disease assessment performed with *P. parasitica* (Figure 18) and indicate that *nim1* alleles can cause resistance or susceptibility.

The samples obtained above were used to analyze *NIM1* gene expression (Figure 17). In wildtype plants *NIM1* mRNA was present in the untreated control samples. Following treatment with SA, INA, BTH or infection with a compatible pathogen the *NIM1* mRNA accumulated to higher levels. Differences in *NIM1* message (mRNA) abundance were observed in the *nim1* alleles compared to wildtype. The abundance of *NIM1* mRNA in untreated mutant plants was lower than observed in the wildtype with the exception of *nim1-2* and *-5* where the amounts were similar. The *nim1-1*, *-3* and *-4* had low levels of *NIM1* message while the *nim1-6* had very low accumulation of *NIM1* mRNA. Increases in *NIM1* mRNA following SA, INA or BTH were observed in *nim1-1*, *-2*, *-3* but not *nim1-5* or *-6*. However, this increase was less than observed in wildtype plants. Following pathogen infection additional bands hybridizing to the *NIM1* cDNA probe were observed in both wildtype and mutants and the *NIM1* mRNA level was elevated relative to untreated controls, except in *nim1-6*.

Figure 18 shows the disease resistance assessment via infection rating of the various *nim1* alleles as well as the NahG plants at various times after innoculation with *Peronospora parasitica*. WsWT indicates the Ws wild type parent line in which the *nim1* alleles are found. The various *nim1* alleles are indicated in the table and the *NahG* plant is indicated also. The NahG plant has been previously published. (Delaney et al. Science 266, pp. 1247-1250 (1994)). The NahG *Arabidopsis* is also described in WO 95/19443.

The *NahG* gene is a gene from *Pseudomonas putida* that converts salicylic acid to catechol, thereby eliminating the accumulation of salicylic acid, a necessary signal transduction component for SAR in plants. Thus, NahG *Arabidopsis* plants do not display normal SAR. In addition, they show much greater susceptibility in general to pathogens. Therefore, the NahG plants serve as a kind of universal susceptibility control. In addition, the NahG plants still respond to the chemical inducers INA and BTH; this is shown in the bottom two panels of the Figure 17.

From Figure 18 it can be seen that the *nim1-4* and *nim1-6* alleles seem to be the most severe; this is most easily observable at the earlier time points, described earlier in the results section herein, and from the results set forth in the EmWa BTH panel, the lowest panel, in the Figure. In addition, the *nim1-5* allele shows the greatest response to both INA and BTH and therefore it is the weakest *nim1* allele.

The NahG plants show very good response to both INA and BTH and look very similar to the *nim1-5* allele. However, at late time points, Day 11 in the Figure, the disease resistance induced in the NahG plants begins to fade, and there is a profound difference between INA and BTH in that the INA-induced resistance fades much faster and more severely than the resistance induced in the NahG plants by BTH. Also seen in these experiments is that INA and BTH induced very good resistance in Ws to EmWa, and the *nim1-1*, *nim1-2* and other *nim1* alleles show virtually no response to SA or INA in regard to disease resistance.

Figure 18 lists the percent of plants that are showing sporulation after infection with the EmWa race of *P. parasitica*, and each of the bar graphs indicates the number of days after infection that the disease resistance was rated.

Northern analysis analyzing the expression of the SAR gene PR1 was also performed on the same samples, as shown in Figure 17. Figure 17 shows that the wild type

plant shows very good response to salicylate, INA, BTH and also to pathogen infection, as manifested by enhanced PR1 gene expression. The *nim1-1* allele, on the other hand, shows only very limited response to all the chemical inducers including pathogen.

The pathogen induction is at least several fold lower in the *nim1-1* allele than it is in the wild type. The *nim1-2*, *nim1-3* and *nim1-6* alleles show response similar to the *nim1-1* allele to the various treatments. However, the *nim1-4* allele shows virtually no expression in response to any of the inducers used. Basically, background level is all that is observed. The *nim1-5* allele shows a very high background level relative to controls with water and that background level is maintained in all the treatments; however, there is limited or no induction by the chemical inducers.

The NahG plants serve as a good control, showing that they are unable to induce PR-1 in the presence of SA; on the other hand, INA and BTH both induce very strong high level expression of PR-1. The effect of pathogen infection is similar to that of SA; there is no expression of PR-1 in the EmWa-treated NahG plants.

These same RNA samples produced in the induction studies were also probed with a *NIM1* gene using a full-length cDNA clone as probed. In Figure 16 it can be seen that INA induces the *NIM1* gene in the wild type Ws allele. However, the *nim1-1* mutation allele shows a lower basal level expression of the *NIM1* gene, and it is not inducible by INA. This is similar to what is observed in the *nim1-3* allele and the *nim1-6* allele. The *nim1-2* allele shows approximately normal levels in the untreated sample and shows similar induction to that of the wild type sample, as does the *nim1-4* allele. The *nim1-5* allele seems to show higher basal level expression of the *NIM1* gene and much stronger expression when induced by chemical inducers.

The induction of *NIM1* by chemical inducers of resistance and other inducers is consistent with its role in pathogen defense and is also further evidence that we have obtained the right gene in the 9.9 kb region.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Novartis AG
(B) STREET: Schwarzwaldallee 215
(C) CITY: Basel
(E) COUNTRY: Switzerland
(F) POSTAL CODE (ZIP): 4002
(G) TELEPHONE: +41 61 69 11 11
(H) TELEFAX: + 41 61 696 79 76
(I) TELEX: 962 991

(ii) TITLE OF INVENTION: GENE CONFERRING DISEASE RESISTANCE IN PLANTS AND USES THEREOF

(iii) NUMBER OF SEQUENCES: 11

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9919 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TGATCATGAA TTGCGTGTAG GGTTGTGTTT TAAAGATAGG GATGAGCTGA AGAAGGCGGT 60

GGACTGGTGT TCCATTAGAG GGCAGCAAAA GTGTGTAGTA CAAGAGATTG AGAAGGACGA 120

GTATACGTTT AAATGCATCA GATGGAAATG CAATTGGTCG CGTCGGGCAG ATTGAATAGA 180

AGAACATGGA CTTGTTAAGA TAACTAAGTG TAGTTGGTCC ACATACTTGT TGTTCTATTA 240

AGCCGGAAAA CTTCAACTTG TAATTTGCAG CAGAAGAGAT TGAGTGTCTG ATCAGGGTAC 300

AACCCACTCT AACAGCAGAG TTGAAAAGTT TGGTGACATG CTTAAAACTT CAAAGCTGCG 360

GGCAGCAGAA CAGGAAGTAA TCAAAGATCA GAGTTTCAGA GTATTGCCTA AACTAATTGG 420

CTGCATTTCA CTCATCTAAT GGGCTACTTG TGGACTGCAA TATGAGCTTT TCCCTAATCC 480

TGAATTTGCA TCCTTCGGTG GCGCGTTTTG GGCGTTTCCA CAGTCCATTG AAGGGTTTCA 540

ACACTGTAGA CCTCTGATCA TAGTGGATTC AAAAGACTTG AACGGCAAGT ACCCTATGAA 600

ATTGATGATT TCCTCAGGAC TCGACGCTGA TGATTGCTTT TTCCCGCTTG CCTTTCCGCT 660

TACCAAAGAA GTGTCCACTG ATAGTTGGCG TTGGTTTCTC ACTAATATCA GAGAGAAGGT 720

AACACAAAGG AAAGACGTTT GCCTCGTCTC CAGTCCTCAC CCGGACATAG TTGCTGTTAT 780

TAACGAACCC GGATCACTGT GGCAAGAACC TTGGGTCTAT CACAGGTTCT GTCTGGATTG 840

TTTTTGCTTA CAATTCCATG ATATTTTTGG AGACTACAAC CTGGTGAGCC TTGTGAAGCA 900

GGCTGGATCC ACAAGTCAGA AGGAAGAATT TGATTCCTAC ATAAAGGACA TCAAAAAGAA 960

GGACTCAGAA GCTCGGAAAT GGTTAGCCCA ATTCCCTCAA AATCAGTGGG CTCTGGCTCA 1020

TGACCAGTGG TCGGAGATAT GGAGTCATGA CGATAGAAAC AGAAGATTTG AGGGCAATTT 1080

GTGAAAGCTT TCAGTCTCTT GGTCTATCAG TGACAGCGAA CGCACCTGCA CATGTGGGAA 1140

GTTTCAATCG AAGAAGTTTC CATGTATGCA CCCAGAAATG GTGCAAAGGA TTGTTAACTT 1200

GTGTCATTCA CAAATGTTGG ATGCAATGGA GCTGACTAGG AGAATGCACC TTACACGCCC 1260

ACTCAGTGTT CTCTTATCTC TAGACCTGAA ACTAACTTGC TGTGTAATTC GAGTTACAAA 1320

AGGTTAAAGG AAGAATTAGG AAGATACATA TAACATGAAT GTTGCCAGAA GTTCAGGGAA 1380

CTTGAATATT CTTTTGGTTC TTGGTGGAAA ATATCCAACA GATGAACAAT TTGACATTAT 1440

TTCACACTTT GATTCTAGCA ACTCTGTAAC ACCATCATGG GTTATTGTTG ATGTACATAA 1500

ATATATATTA CAAATCTGTA TACCATTGGT TCAAATTGTT ACAACATTTG TTTGAAGCAC 1560

ACCTGCAGCA ATAATACACA GGATGCAAAA CGAAGAGCGA AACTATATGA CGCCAACGAT 1620

AGACATAAAC AGTTACAGTC ATCATGAAAA CAGAATTATA TGGTACAGCA AAAATTACAC 1680

TAAGAGGCAA GAGTCTCACC GACGACGATG AGAGAGTTTA CGGTTAGACC TCTTTCCACC 1740

GGTTGATTTC GATGTGGAAG AAGTCGAATC TGTCAGGGAC GAATTTCCTA ATTCCAAATT 1800

GTCCTCACTA AAGGCCTTCT TTAGTGTCTC TTGTATTTCC ATGTACCTTT GCTTCTTTTG 1860

TAGTCGTTTC TCAGCAGTGT CGTCTTCTCC GCAAGCCAGT TGAGTCAAGT CCTCACAGTT 1920

CATAATCTGG TCGAGCACTG CCGAACAGCG CGGGAAGAAT CGTTTCCCGA GTTCCACTGA 1980

TGATAAAAAA AACAAGGTCA GACAGCAAGT AACAAAACCA TGTTTAAAGA TCATTTAGTT 2040

TTGTTTTTTG TGATAAGGAG TCCGATGAAG TGGGTGAGAA TCCATACCGG TTTTAGAAAG 2100

CGCTTTTAGT CTACTTTGAT GCTCTTCTAG GATTCTGAAA GGTGCTATCT TTACACCCGG 2160

TGATGTTCTC TTCGTACCAG TGAGACGGTC AGGCTCGAGG CTAGTCACTA TGAACTCACA 2220

TGTTCCCTTC ATTTCGGCGA TCTCCATTGC AGCTTGTGCT TCCGTTGGAA AAAGACGTTG 2280

AGCAAGTGCA ACTAAACAGT GGACGACACA AAGAATAGTT ATCATTAGTT CACTCAGTTT 2340

CCTAATAGAG AGGACATAAA TTTAATTCAA ACATATAAGA AATAAGACTT GATAGATACC 2400

TCTATTTTCA AGATCGAGCA GCGTCATCTT CAATTCATCG GCCGCCACTG CAAAAGAGGG 2460

AGGAACATCT CTAGGAATTT GTTCTCGTTT GTCTTCTTGC TCTAGTATTT CTACACATAG 2520

TCGGCCTTTG AGAGAATGCT TGCATTGCTC CGGGATATTA TTACATTCAA CCGCCATAGT 2580

GGCTTGTTTT GCGATCATGA GTGCGGTTCT ACCTTCCAAA GTTGCTTCTG ATGCACTTGC 2640

ACCTTTTTCC AATAGAGATA GTATCAATTG TGGCTCCTTC CGCATCGCAG CAACATGAAG 2700

CACCGTATAT CCCCTCGGAT TCCTATGGTT GACATCGGCA AGATCAAGTT TTAAAAGATC 2760

TGTTGCGGTC TTCACATTGC AATATGCAAC AGCGAAATGA AGAGCACACG CATCATCTAG 2820

ATTGGTGTGA TCCTCTTTCA AAAGCAACTT GACTAACTCA ATATCATCCG AGTCAAGTGC 2880

CTTATGTACA TTCGAGACAT GTTTCTTTAC TTTAGGTACC TCCAAACCAA GCTCTTTACG 2940

TCTATCAATT ATCTCTTTAA CAAGCTCTTC CGGCAATGAC TTTTCAAGAC TAACCATATC 3000

TACATTAGAC TTGACAATAA TCTCTTTACA TCTATCCAAT AGCTTCATAC AAGCTTTACC 3060

ACATATATTA GCAAGCTTGA GTATAACCAA TGTGTCCTCT ATAACAACTT TGTCTACAAC 3120

GTCCAATAAG TGCCTCTGAA ATACAAATAC AAGTACTCAA GTAAGAACAT ATTCATGAAT 3180

GTGTAACCAT AGCTTAATGC AGATGGTGTT TTACCTGATA GAGAGTAATT AATTCAGGGA 3240

TCTTGAAGAT GAAAGCCAAA TAGAGAACCT CCAACATGAA ATCCACCGCC GGCCGGCAAG 3300

CCACGTGGCA GCAATTCTCG TCTGCGCATT CAGAAACTCC TTTAGGCGGC GGTCTCACTC 3360

TGCTGCTGTA AACATAAGCC AAAACAGTCA CAACCGAATC GAAACCGACT TCGTAATCCT 3420

TGGCAATCTC CTTAAGCTCG AGCTTCACGG CGGCGGTGTT GTTGGAGTCT TTCTCCTTCT 3480

TAGCGGCGGC TAAAGCGCTC TTGAAGAAAG AGCTTCTCGC TGACAAAACG CACCGGTGGA 3540

AAGAAACTTC CCGGCCGTCG GAGAGAACAA GCTTAGCGTC GCTGTAGAAA TCATCCGGCG 3600

AGTCAAAGAC GGATTCGAAG CTGTTGGAGA GCAATTGCAG AGCAGATACA TCAGGTCCGG 3660

TGAGTACTTG TTCGGCGGCC AGATAAACAA TAGAGGAGTC GGTGTTATCG GTAGCGACGA 3720

AACTAGTGCT GCTGATTTCA TAAGAATCGG CGAATCCATC AATGGTGGTG TCCATCAACA 3780

GGTTCCGATG AATTGAAATT CACAAATTAA AGAGATCTCT GCTAATCAAC GAAGAGACCT 3840

TATCAACTGG ATTTGGTTAA AGATCGAAGA TAACCATTGA CGAGCAGAGC CAAGTCAAGT 3900

CAACGAGAGT GGTGGTGAGA TATGAAGAAG CATCCTCGTC CCACGGTTTA CATTTCACCA 3960

AAACCGGTAA ATTTCCAGGA AAGGAATCTT TGTCAGAGAT CTTTTTTAAA AAGATATAAC 4020

AGGAAGCTAA ACCGGTTCGG GTTATAAATG TTAGTATTTA TACCGGAGAC ATTTTGTGTT 4080

GCTAATTTTT GTATATGAGA AGTTCAATCC GGTTCGGTAA GCCCCTGAAC CAAACTAGAT 4140

TTGGAGATGA TATAAATATA TAAAATTTAT TTTTCATCCG GTTCGTTATT TTCATATAAA 4200

TATATAAATA TTATTTTTTA AATTTAAGAA TTAGATTTAC ATGTGAAAGT TACATTTCTG 4260

TTTATTTTCT TTGAAGTAAA ATGATAAAGG GAACGTATAT TAAGTTTCAT GCTTTATTCA 4320

CATAAGTTTT GTAATGTATA TTATATTTTT CGTTTATTGA AAAAGTAATT TTCAGTGTTC 4380

AGCATGTTTA CACTATAATT AAATCAAGTC GAATATTTCC TGGAACTATT CTCCTTGTTC 4440

TATAGCAAAT GAAAACGCTC TTCACAACAA AATCATTATA GATATAGGAA TAAATTACAT 4500

TAAAAACATG AAAGTCATAA TGAATATATT TTTTTAATTA GGATTTGATT TAAAAACAAT 4560

TATTGTATAC ATATAAAAGA CTTCTTTAGT TATTTGCCTT CAACTTCTCG TTCTGAATCA 4620

TGCGATAAAT CAGCTTTTTC AATAACTACG ACGTAAAAGC AAATTCATAA CACGTCTAAA 4680

CAAATTTGGC TCATCCTTCA CTTGATTGGT GTTTTCCGGA CTCGATGTTG CTGGAAACTG 4740

AGAAGAAGAA GGAATCTGCA TAATCACCTC TTGGTTCCTC ACCGGTAGAC TCATTTTGTT 4800

GGATCGAAAA CGATCGAGAT CAGAAAATGA AAAGATAGGT TAAAGATGCC TATGAATACA 4860

ACAACGTAAG ATTATGTTGA ATAAACAGAG TACTTTATAT AGGAGTTATA ATAAGGTAAA 4920

TAAATTATTG CTTTCCGCGT TTTTTACTTT TGTATTTCTT AAATGATAAG TTAAATTAGG 4980

ATAAGATTTG TATGATTTTA AGTAAATTTA CAATAACTCT CTATAACTCA ATAGCATCAC 5040

ATATTTAATT AATTTTACTA ATTATCTTTT GAACAATTTT ATGAAATAGT TTTCTTTTAA 5100

TTAATTTTTT AAAATGATAT ATTATAAAAT TTAATTGAAT CAATCTGATA TAATTTTTTT 5160

ATCTTCTACC ATCTATTATA GTTGATAAAT ATTGTGATAA ACTTTAGATA AACACCCAAT 5220

TGCCAAATAT TTAATAAATT TTGTGTACCA TGCGTTTTTT TTGGAGAATA TATATACGTG 5280

GACAGCATAC CGTACATATA TTGTATAAAA GCTTATAAAA CATAGATACG GGTTATATTG 5340

GTAAGCTATA AATATATGTA AACAATAGTA AGATATTACG TGTTGTGTCT AAATATGTGT 5400

TGCTTTAGAT ATTATGTATA TCTAATATAT TAAAATATCT TTTATTAACT AATATATTAT 5460

TTAAGAGAGA AAATTGGGAC ACTATTTTCT ATACAGTAAC TGTTTTCAAC TATAAACAGG 5520

AACCCTTGAT ATAATAAAAT AACTAGCCAA AAAATCAGAT TAAATATTCA TAAAACAATG 5580

TTTGGTATTA TTACATAAAC CTAAGAAACA AAATTCAATA TTCCTTTTTA CCTTATAAAA 5640

AACAATTAAA CATCACTAGA TATATTTATG CCCCACAATG AGCGAGCCAA TTGAGACTTG 5700

AGACTTGAGA TCCTTGTCAA CTACGTTTGC ATTTGTCGGC CCATTTTTTT TATTTTTTTT 5760

TTAAAGTGTC GGCCCGTTGC TTCTTCCGTT CAGATCAACC CTCTCGTAAT CAGAACAAAA 5820

CGGAAAACAA ACGAAAGAAC AATCAGATCC CTCTTTTTTT GCATAAACTA AATTCAACTT 5880

CTCTGCGTTT ATGTTGTAGA GGCAACCACG ATCACTACTA CGAAACAATA CAACGTCGTT 5940

GCTTGGAGTC CACGTAATCA AATCTACTCC AATGCTTTTA ATATCTTTCA CTTTAACCCA 6000

CGACTTTTCA AAACTGCTCT TTAAAACCCA TAACTCGTGA ACATCTTCTT GATCTTTGTT 6060

TGTCCACTGA CGAATAGCAC CTAGCTTCCC TTCGTATCTG ACTAATCCTG AGAAAACATC 6120

AGAGTTCGGA GTATGGAAGA AGGACCAAGT TTCGGTTTTG AGACAAAACC GGATCACATT 6180

GTTGTTCCGT GATATCCAAT GCAAGAACCC CGAAACTTGT ATCGGGTTGG AAAAAATTAA 6240

TCTGTCTGTT TTTGGTAGAC GCAAATTTTC TAATCTCTTC CAGGTAAACG AATCAGAATC 6300

GAAAACTTCG CACATAAAAG TTCTGTGATT CAAATGGTAG ATACCCCGAG ACATACACAT 6360

ACGCCGAGAC TGCGAAAGCC TTTGTATTTT ATACCGGAAA GGGTTCAATC CGATTACCGC 6420

TAAACCCAAT GACATATCCC AACCCTTCAC TTCTGGCTTT GGTATGACCT GATACTGTTT 6480

AGTGGTTGGT TTGAAGACTA TGTATCCACG TGATGGTTTT GTATACTTAA CACAAAGCAA 6540

TATCCCATGA CTTGCATCAC AAGCTTCGAT CTTTATCATT CCGGGTGGCA GAAAGTCGAT 6600

GGAGACTCCA TTGTTTTGTA AATCACTCCT CTCATGGACA AAACTGGTTC GAAGTTCGTG 6660

TCCTTTTACT ATGTAGTGTT GTATGAAGTA TCCCGAAATA CGATTGGTTC TAAGGAGATT 6720

AAGATTGACA AACCATGACT CGTAGCTTCT CTTGTTGCAC TCTTTATTCA GGAGCCTGAA 6780

TTTTCCGATT TTTGACGCCG GAAGATAAGA AAGAAATTCT TGGATCATGT CTTGATTTAT 6840

CACCGGAGAA CTCATGATCC TGTCGGGAAT AAAGAGATGA GCACGATCAC TGAATGAGAA 6900

ATGAAAAAAT CAGGATCGGT AGAGAACAAC TTATGATGAA TAAAGTGTTT ATATATCCTT 6960

TCTTTGTTTA AGGAAAGTAT CAAAATTTGC CTTTTTCTTC GCTAGTCCTA AAACAAACAA 7020

ATTAACCAAA AGATAAAATC TTTCATGATT AATGTTACTT GTGATACCTT AAGCCAAAAC 7080

TTTATCTTTA GACTTTTAAC CAAATCTACA GTAATTTAAT TGCTAGACTT AGGAAACAAC 7140

TTTTTTTTTT ACCCAACAAT CTTTGGATTT TAATTGTTTT TTTTTCTACT AATAGATTAA 7200

CAACTCATTA TATAATAATG TTTCTATCAT AATTGACAAT TCTTTCTTTT TAATAAACAT 7260

CCAGCTTGTA TAATAATCCA CAAGTCAATT TCACCATTTT GGCCAATTTA TTTTCTTATA 7320

AAAATTAGCA CAAAAAAGAT TATCATTGTT TAGCAGATTT AATTTCTAAT TAACTTACGT 7380

AATTTCCATT TTCCATAGAT TTATCTTTCT TTTTATTTCC TTAGTTATCT TAGTACTTTC 7440

TTAGTTTCCT TAGTAATTTT AAATTTTAAG ATAATATATT GAAATTAAAA GAAGAAAAAA 7500

AACTCTAGTT ATACTTTTGT TAAATGTTTC ATCACACTAA CTAATAATTT TTTTTAGTTA 7560

AATTACAATA TATAAACACT GAAGAAAGTT TTTGGCCCAC ACTTTTTTGG GATCAATTAG 7620

TACTATAGTT AGGGGAAGAT TCTGATTTAA AGGATACCAA AAATGACTAG TTAGGACATG 7680

AATGAAAACT TATAATCTCA ATAACATACA TACGTGTTAC TGAACAATAG TAACATCTTA 7740

CGTGTTTTGT CCATATATTT GTTGCTTATA AATATATTCA TATAACAATG TTTGCATTAA 7800

GCTTTTAAGA AGCACAAAAC CATATAACAA AATTAAATAT TCCTATCCCT ACCAAAAAAA 7860

AAAATTAAAT ATTCCTACAG CCTTGTTGAT TATTTTATGC CCTACGTTGA GCCTTGTTGA 7920

CTAGTTTGCA TTTGTCGGTC CATTTCTTCT TCCGTCCAGA TCAACCCTCT CGTAATCAGA 7980

ACAAAAGGGG AAACAAACGT AAGAGGCAAA ATCCTTGTTT GTATGAACTA AGTTTAACTT 8040

CTCTGTGTTT AAGTTGTAGA GGCAAACATG ATCCCAACTA GAAAGCATTA CGACGTCGTT 8100

GCTTGGTATC CACGTAATAT GCTCTACTCC AATGCTTTCA ATATCTTTCA CTTTTTCCCA 8160

CGACTTTTCA AAACTGCTCT TTAAAACCCA TAATCTGTGA ACATCTTCTT GATTGTTGTT 8220

TATCCAGTGA CGAATAACAC CTAGCTTCCC TTCGTAGCTG ACTAACTCTG GGAATAAACC 8280

AACGTTTGGA GTATGTAAGA AAGACCAAGT TTCGGTTTTG GGACATAACC GGATCACATT 8340

GTGGTTCCAT GATCTCCAAT GCAAGAACCC TGAAGCTTGT ACCGGGTTTG AAAGAATTAG 8400

ACCGTCTGTT CTCGGTAGAC GCAAATTTTT TAATCTCTTC CACATAAACG AATCGGAATC 8460

AAAAACTTCG CACGCAAAAG TTCTGAGATT CCGAGTCATA CCAGGCGATT TCGAAAGCCT 8520

AAATATTTTA TACCGGAAAG GCTGCAATCC GGTTACCGTT AGACCTAATG ACTTATCACA 8580

ACTCCTCACT TTTGGGTTTG GTATGATCTG ATACTGTTTT GTTGTTGGTT TGCAGACTAT 8640

GTATTCCGGT ATTGGTCTTG TATCATTATA ACAAAGCAAT ATCCCATGAC GTGCATCACA 8700

AGCTTTGATC TTTACCTCTC CTTGTGGCAG AAAATCGATG GAGACTCCTT TGTTATCCAA 8760

ATCTCTCCTC TCATGGAAAA AACTGGTATC AAGTTTGTAT CCTCTTTCGT AGCGTTCTAG 8820

GAAGTATCCA GAGATATTGT TGGTTCGATG GAGATTTAGG TTGACAAACC AAGACTCGTA 8880

GCTTCTCTTG TTGCACTCTT TATTGATGAG CCTCAATTTT CCGATTTCGG ACCCCCGAAG 8940

ATAAGAAAGA ACCTCTTGGA TCGTGTCCTG ATTTATCACC GGAGAACTCA TGATCTTATT 9000

GGAAAAAAGA AAGAAAGAGA TGAGCACGAT CAGTGAATGA GATATATAGA AATCAGGATT 9060

GGTAGAGAAC CGACGATGAT GAATATACAA GTGTTTATAA GTATCACAAA TTGCCTTTTT 9120

CTTCGCTAGT CCCAAAACAA GCAAATTAAC CAAAGATAAA ATCTTCATTA ATGTTTTCCT 9180

TTTTCTTCGC CAGTCCCAGA TAAAAATATA TATAAAATAT TTCATTAGGT TACTTGTAGT 9240

ACCTTGAGCC CAAAGTTTCT CTTTTGACTT TTAACCAAAT TAACAGTAAT TTAATAGCTA 9300

GACTTAGAAA ACAACATTTT GTATATATAT TCTTTGACAT CAAAATTCAA CAATCTTTGG 9360

GTTTCTATAG TGTTTTTTTT CTTATTCTAA TAGATTACCA CTCATTATAT CATATACAAA 9420

GTGTTTCCTT TTCAATCAAC ATCCATTTTC TTTAAAAATT AGCAAGTTTG TTCTTATATC 9480

ATCATTCAGC AGATTTCTTA ATTAAACTTA GTGATTTCCA TTTTGCACCT ATATGTTTCT 9540

CTTTCTTAGT TTAGTACTTT AAATTTTCAT ATATATAATT TATTAAAATT AAAAGTAAAA 9600

ACTCCAGTTT AACTTATGTT AAATGTTTCA TCACACTAAA AGAGCATTAA GTAATAAATA 9660

TTTTAGCTTT ATGAAAAAAA ATATCAAATC ACTGAAGACA TTTGTTGGCC TATACTCTAT 9720

TTTTTATTTG GCCAATTAGT AATAGACTAA TAGTAACTCA TATGATATCT CTCTAATTCT 9780

GGCGAAACGA ATATTCTGAT TCTAAAGATA GTAAAAATGA ATTTTGATGA AGGGAATACT 9840

ATTTCACACA CCTAGAAAGA GTAAGGTAGA AACCTTTTTT TTTTTGGTCA GATTCTTGTA 9900

TCAAGAAGTT CTCATCGAT 9919

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5655 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 2787..3347
(D) OTHER INFORMATION: /product= "1st exon of NIM1"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 3427..4162
(D) OTHER INFORMATION: /product= "2nd exon of NIM1"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 4271..4474
(D) OTHER INFORMATION: /product= "3rd exon of NIM1"

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 4586..4866
(D) OTHER INFORMATION: /product= "4th exon of NIM1"

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: join(2787..3347, 3427..4162, 4271..4474, 4586..4866)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TGTGATGCAA GTCATGGGAT ATTGCTTTGT GTTAAGTATA CAAAACCATC ACGTGGATAC 60

ATAGTCTTCA AACCAACCAC TAAACAGTAT CAGGTCATAC CAAAGCCAGA AGTGAAGGGT 120

TGGGATATGT CATTGGGTTT AGCGGTAATC GGATTGAACC CTTTCCGGTA TAAAATACAA 180

AGGCTTTCGC AGTCTCGGCG TATGTGTATG TCTCGGGGTA TCTACCATTT GAATCACAGA 240

ACTTTTATGT GCGAAGTTTT CGATTCTGAT TCGTTTACCT GGAAGAGATT AGAAAATTTG 300

CGTCTACCAA AAACAGACAG ATTAATTTTT TCCAACCCGA TACAAGTTTC GGGGTTCTTG 360

CATTGGATAT CACGGAACAA CAATGTGATC CGGTTTTGTC TCAAAACCGA AACTTGGTCC 420

TTCTTCCATA CTCCGAACTC TGATGTTTTC TCAGGATTAG TCAGATACGA AGGGAAGCTA 480

GGTGCTATTC GTCAGTGGAC AAACAAAGAT CAAGAAGATG TTCACGAGTT ATGGGTTTTA 540

AAGAGCAGTT TTGAAAAGTC GTGGGTTAAA GTGAAAGATA TTAAAAGCAT TGGAGTAGAT 600

TTGATTACGT GGACTCCAAG CAACGACGTT GTATTGTTTC GTAGTAGTGA TCGTGGTTGC 660

CTCTACAACA TAAACGCAGA GAAGTTGAAT TTAGTTTATG CAAAAAAAGA GGGATCTGAT 720

TGTTCTTTCG TTTGTTTTCC GTTTTGTTCT GATTACGAGA GGGTTGATCT GAACGGAAGA 780

AGCAACGGGC CGACACTTTA AAAAAAAAAT AAAAAAAATG GGCCGACAAA TGCAAACGTA 840

GTTGACAAGG ATCTCAAGTC TCAAGTCTCA ATTGGCTCGC TCATTGTGGG GCATAAATAT 900

ATCTAGTGAT GTTTAATTGT TTTTTATAAG GTAAAAAGGA ATATTGAATT TTGTTTCTTA 960

GGTTTATGTA ATAATACCAA ACATTGTTTT ATGAATATTT AATCTGATTT TTTGGCTAGT 1020

TATTTTATTA TATCAAGGGT TCCTGTTTAT AGTTGAAAAC AGTTACTGTA TAGAAAATAG 1080

TGTCCCAATT TTCTCTCTTA AATAATATAT TAGTTAATAA AAGATATTTT AATATATTAG 1140

ATATACATAA TATCTAAAGC AACACATATT TAGACACAAC ACGTAATATC TTACTATTGT 1200

TTACATATAT TTATAGCTTA CCAATATAAC CCGTATCTAT GTTTTATAAG CTTTTATACA 1260

ATATATGTAC GGTATGCTGT CCACGTATAT ATATTCTCCA AAAAAAACGC ATGGTACACA 1320

AAATTTATTA AATATTTGGC AATTGGGTGT TTATCTAAAG TTTATCACAA TATTTATCAA 1380

CTATAATAGA TGGTAGAAGA TAAAAAAATT ATATCAGATT GATTCAATTA AATTTTATAA 1440

TATATCATTT TAAAAAATTA ATTAAAAGAA AACTATTTCA TAAAATTGTT CAAAAGATAA 1500

TTAGTAAAAT TAATTAAATA TGTGATGCTA TTGAGTTATA GAGAGTTATT GTAAATTTAC 1560

TTAAAATCAT ACAAATCTTA TCCTAATTTA ACTTATCATT TAAGAAATAC AAAAGTAAAA 1620

AACGCGGAAA GCAATAATTT ATTTACCTTA TTATAACTCC TATATAAAGT ACTCTGTTTA 1680

TTCAACATAA TCTTACGTTG TTGTATTCAT AGGCATCTTT AACCTATCTT TTCATTTTCT 1740

GATCTCGATC GTTTTCGATC CAACAAAATG AGTCTACCGG TGAGGAACCA AGAGGTGATT 1800

ATGCAGATTC CTTCTTCTTC TCAGTTTCCA GCAACATCGA GTCCGGAAAA CACCAATCAA 1860

GTGAAGGATG AGCCAAATTT GTTTAGACGT GTTATGAATT TGCTTTTACG TCGTAGTTAT 1920

TGAAAAAGCT GATTTATCGC ATGATTCAGA ACGAGAAGTT GAAGGCAAAT AACTAAAGAA 1980

GTCTTTTATA TGTATACAAT AATTGTTTTT AAATCAAATC CTAATTAAAA AAATATATTC 2040

ATTATGACTT TCATGTTTTT AATGTAATTT ATTCCTATAT CTATAATGAT TTTGTTGTGA 2100

AGAGCGTTTT CATTTGCTAT AGAACAAGGA GAATAGTTCC AGGAAATATT CGACTTGATT 2160

TAATTATAGT GTAAACATGC TGAACACTGA AAATTACTTT TTCAATAAAC GAAAAATATA 2220

ATATACATTA CAAAACTTAT GTGAATAAAG CATGAAACTT AATATACGTT CCCTTTATCA 2280

TTTTACTTCA AAGAAAATAA ACAGAAATGT AACTTTCACA TGTAAATCTA ATTCTTAAAT 2340

TTAAAAAATA ATATTTATAT ATTTATATGA AAATAACGAA CCGGATGAAA AATAAATTTT 2400

ATATATTTAT ATCATCTCCA AATCTAGTTT GGTTCAGGGG CTTACCGAAC CGGATTGAAC 2460

TTCTCATATA CAAAAATTAG CAACACAAAA TGTCTCCGGT ATAAATACTA ACATTTATAA 2520

CCCGAACCGG TTTAGCTTCC TGTTATATCT TTTTAAAAAA GATCTCTGAC AAAGATTCCT 2580

TTCCTGGAAA TTTACCGGTT TTGGTGAAAT GTAAACCGTG GGACGAGGAT GCTTCTTCAT 2640

ATCTCACCAC CACTCTCGTT GACTTGACTT GGCTCTGCTC GTCAATGGTT ATCTTCGATC 2700

TTTAACCAAA TCCAGTTGAT AAGGTCTCTT CGTTGATTAG CAGAGATCTC TTTAATTTGT 2760

GAATTTCAAT TCATCGGAAC CTGTTG ATG GAC ACC ACC ATT GAT GGA TTC GCC 2813

```
                                Met Asp Thr Thr Ile Asp Gly Phe Ala
                                  1               5

GAT TCT TAT GAA ATC AGC AGC ACT AGT TTC GTC GCT ACC GAT AAC ACC      2861
Asp Ser Tyr Glu Ile Ser Ser Thr Ser Phe Val Ala Thr Asp Asn Thr
 10                  15                  20                  25

GAC TCC TCT ATT GTT TAT CTG GCC GCC GAA CAA GTA CTC ACC GGA CCT      2909
Asp Ser Ser Ile Val Tyr Leu Ala Ala Glu Gln Val Leu Thr Gly Pro
                 30                  35                  40

GAT GTA TCT GCT CTG CAA TTG CTC TCC AAC AGC TTC GAA TCC GTC TTT      2957
Asp Val Ser Ala Leu Gln Leu Leu Ser Asn Ser Phe Glu Ser Val Phe
             45                  50                  55

GAC TCG CCG GAT GAT TTC TAC AGC GAC GCT AAG CTT GTT CTC TCC GAC      3005
Asp Ser Pro Asp Asp Phe Tyr Ser Asp Ala Lys Leu Val Leu Ser Asp
         60                  65                  70

GGC CGG GAA GTT TCT TTC CAC CGG TGC GTT TTG TCA GCG AGA AGC TCT      3053
Gly Arg Glu Val Ser Phe His Arg Cys Val Leu Ser Ala Arg Ser Ser
     75                  80                  85

TTC TTC AAG AGC GCT TTA GCC GCC GCT AAG AAG GAG AAA GAC TCC AAC      3101
Phe Phe Lys Ser Ala Leu Ala Ala Ala Lys Lys Glu Lys Asp Ser Asn
 90                  95                  100                 105

AAC ACC GCC GCC GTG AAG CTC GAG CTT AAG GAG ATT GCC AAG GAT TAC      3149
Asn Thr Ala Ala Val Lys Leu Glu Leu Lys Glu Ile Ala Lys Asp Tyr
                110                 115                 120

GAA GTC GGT TTC GAT TCG GTT GTG ACT GTT TTG GCT TAT GTT TAC AGC      3197
Glu Val Gly Phe Asp Ser Val Val Thr Val Leu Ala Tyr Val Tyr Ser
            125                 130                 135

AGC AGA GTG AGA CCG CCG CCT AAA GGA GTT TCT GAA TGC GCA GAC GAG      3245
Ser Arg Val Arg Pro Pro Pro Lys Gly Val Ser Glu Cys Ala Asp Glu
        140                 145                 150

AAT TGC TGC CAC GTG GCT TGC CGG CCG GCG GTG GAT TTC ATG TTG GAG      3293
Asn Cys Cys His Val Ala Cys Arg Pro Ala Val Asp Phe Met Leu Glu
    155                 160                 165
```

```
GTT CTC TAT TTG GCT TTC ATC TTC AAG ATC CCT GAA TTA ATT ACT CTC      3341
Val Leu Tyr Leu Ala Phe Ile Phe Lys Ile Pro Glu Leu Ile Thr Leu
170                 175                 180                 185

TAT CAG GTAAAACACC ATCTGCATTA AGCTATGGTT ACACATTCAT GAATATGTTC      3397
Tyr Gln

TTACTTGAGT ACTTGTATTT GTATTTCAG AGG CAC TTA TTG GAC GTT GTA GAC     3450
                               Arg His Leu Leu Asp Val Val Asp
                                       190                 195

AAA GTT GTT ATA GAG GAC ACA TTG GTT ATA CTC AAG CTT GCT AAT ATA     3498
Lys Val Val Ile Glu Asp Thr Leu Val Ile Leu Lys Leu Ala Asn Ile
                200                 205                 210

TGT GGT AAA GCT TGT ATG AAG CTA TTG GAT AGA TGT AAA GAG ATT ATT     3546
Cys Gly Lys Ala Cys Met Lys Leu Leu Asp Arg Cys Lys Glu Ile Ile
            215                 220                 225

GTC AAG TCT AAT GTA GAT ATG GTT AGT CTT GAA AAG TCA TTG CCG GAA     3594
Val Lys Ser Asn Val Asp Met Val Ser Leu Glu Lys Ser Leu Pro Glu
        230                 235                 240

GAG CTT GTT AAA GAG ATA ATT GAT AGA CGT AAA GAG CTT GGT TTG GAG     3642
Glu Leu Val Lys Glu Ile Ile Asp Arg Arg Lys Glu Leu Gly Leu Glu
    245                 250                 255

GTA CCT AAA GTA AAG AAA CAT GTC TCG AAT GTA CAT AAG GCA CTT GAC     3690
Val Pro Lys Val Lys Lys His Val Ser Asn Val His Lys Ala Leu Asp
260                 265                 270                 275

TCG GAT GAT ATT GAG TTA GTC AAG TTG CTT TTG AAA GAG GAT CAC ACC     3738
Ser Asp Asp Ile Glu Leu Val Lys Leu Leu Leu Lys Glu Asp His Thr
                280                 285                 290

AAT CTA GAT GAT GCG TGT GCT CTT CAT TTC GCT GTT GCA TAT TGC AAT     3786
Asn Leu Asp Asp Ala Cys Ala Leu His Phe Ala Val Ala Tyr Cys Asn
            295                 300                 305

GTG AAG ACC GCA ACA GAT CTT TTA AAA CTT GAT CTT GCC GAT GTC AAC     3834
Val Lys Thr Ala Thr Asp Leu Leu Lys Leu Asp Leu Ala Asp Val Asn
        310                 315                 320
```

```
CAT AGG AAT CCG AGG GGA TAT ACG GTG CTT CAT GTT GCT GCG ATG CGG      3882
His Arg Asn Pro Arg Gly Tyr Thr Val Leu His Val Ala Ala Met Arg
    325                 330                 335

AAG GAG CCA CAA TTG ATA CTA TCT CTA TTG GAA AAA GGT GCA AGT GCA      3930
Lys Glu Pro Gln Leu Ile Leu Ser Leu Leu Glu Lys Gly Ala Ser Ala
340                 345                 350                 355

TCA GAA GCA ACT TTG GAA GGT AGA ACC GCA CTC ATG ATC GCA AAA CAA      3978
Ser Glu Ala Thr Leu Glu Gly Arg Thr Ala Leu Met Ile Ala Lys Gln
                360                 365                 370

GCC ACT ATG GCG GTT GAA TGT AAT AAT ATC CCG GAG CAA TGC AAG CAT      4026
Ala Thr Met Ala Val Glu Cys Asn Asn Ile Pro Glu Gln Cys Lys His
            375                 380                 385

TCT CTC AAA GGC CGA CTA TGT GTA GAA ATA CTA GAG CAA GAA GAC AAA      4074
Ser Leu Lys Gly Arg Leu Cys Val Glu Ile Leu Glu Gln Glu Asp Lys
        390                 395                 400

CGA GAA CAA ATT CCT AGA GAT GTT CCT CCC TCT TTT GCA GTG GCG GCC      4122
Arg Glu Gln Ile Pro Arg Asp Val Pro Pro Ser Phe Ala Val Ala Ala
    405                 410                 415

GAT GAA TTG AAG ATG ACG CTG CTC GAT CTT GAA AAT AGA G                4162
Asp Glu Leu Lys Met Thr Leu Leu Asp Leu Glu Asn Arg
420                 425                 430

GTATCTATCA AGTCTTATTT CTTATATGTT TGAATTAAAT TTATGTCCTC TCTATTAGGA    4222

AACTGAGTGA ACTAATGATA ACTATTCTTT GTGTCGTCCA CTGTTTAG  TT GCA CTT     4278
                                                      Val Ala Leu
                                                              435

GCT CAA CGT CTT TTT CCA ACG GAA GCA CAA GCT GCA ATG GAG ATC GCC      4326
Ala Gln Arg Leu Phe Pro Thr Glu Ala Gln Ala Ala Met Glu Ile Ala
                440                 445                 450

GAA ATG AAG GGA ACA TGT GAG TTC ATA GTG ACT AGC CTC GAG CCT GAC      4374
Glu Met Lys Gly Thr Cys Glu Phe Ile Val Thr Ser Leu Glu Pro Asp
            455                 460                 465
```

```
CGT CTC ACT GGT ACG AAG AGA ACA TCA CCG GGT GTA AAG ATA GCA CCT      4422
Arg Leu Thr Gly Thr Lys Arg Thr Ser Pro Gly Val Lys Ile Ala Pro
        470                 475                 480

TTC AGA ATC CTA GAA GAG CAT CAA AGT AGA CTA AAA GCG CTT TCT AAA      4470
Phe Arg Ile Leu Glu Glu His Gln Ser Arg Leu Lys Ala Leu Ser Lys
    485                 490                 495

ACC G GTATGGATTC TCACCCACTT CATCGGACTC CTTATCACAA AAAACAAAAC         4524
Thr
500

TAAATGATCT TTAAACATGG TTTTGTTACT TGCTGTCTGA CCTTGTTTTT TTTATCATCA    4584

G  TG GAA CTC GGG AAA CGA TTC TTC CCG CGC TGT TCG GCA GTG CTC        4629
  Val Glu Leu Gly Lys Arg Phe Phe Pro Arg Cys Ser Ala Val Leu
                  505                 510                 515

GAC CAG ATT ATG AAC TGT GAG GAC TTG ACT CAA CTG GCT TGC GGA GAA      4677
Asp Gln Ile Met Asn Cys Glu Asp Leu Thr Gln Leu Ala Cys Gly Glu
                520                 525                 530

GAC GAC ACT GCT GAG AAA CGA CTA CAA AAG AAG CAA AGG TAC ATG GAA      4725
Asp Asp Thr Ala Glu Lys Arg Leu Gln Lys Lys Gln Arg Tyr Met Glu
            535                 540                 545

ATA CAA GAG ACA CTA AAG AAG GCC TTT AGT GAG GAC AAT TTG GAA TTA      4773
Ile Gln Glu Thr Leu Lys Lys Ala Phe Ser Glu Asp Asn Leu Glu Leu
        550                 555                 560

GGA AAT TCG TCC CTG ACA GAT TCG ACT TCT TCC ACA TCG AAA TCA ACC      4821
Gly Asn Ser Ser Leu Thr Asp Ser Thr Ser Ser Thr Ser Lys Ser Thr
    565                 570                 575

GGT GGA AAG AGG TCT AAC CGT AAA CTC TCT CAT CGT CGT CGG TGA          4866
Gly Gly Lys Arg Ser Asn Arg Lys Leu Ser His Arg Arg Arg  *
580                 585                 590

GACTCTTGCC TCTTAGTGTA ATTTTTGCTG TACCATATAA TTCTGTTTTC ATGATGACTG    4926

TAACTGTTTA TGTCTATCGT TGGCGTCATA TAGTTTCGCT CTTCGTTTTG CATCCTGTGT    4986

ATTATTGCTG CAGGTGTGCT TCAAACAAAT GTTGTAACAA TTTGAACCAA TGGTATACAG    5046
```

ATTTGTAATA TATATTTATG TACATCAACA ATAACCCATG ATGGTGTTAC AGAGTTGCTA 5106

GAATCAAAGT GTGAAATAAT GTCAAATTGT TCATCTGTTG GATATTTTCC ACCAAGAACC 5166

AAAAGAATAT TCAAGTTCCC TGAACTTCTG GCAACATTCA TGTTATATGT ATCTTCCTAA 5226

TTCTTCCTTT AACCTTTTGT AACTCGAATT ACACAGCAAG TTAGTTTCAG GTCTAGAGAT 5286

AAGAGAACAC TGAGTGGGCG TGTAAGGTGC ATTCTCCTAG TCAGCTCCAT TGCATCCAAC 5346

ATTTGTGAAT GACACAAGTT AACAATCCTT TGCACCATTT CTGGGTGCAT ACATGGAAAC 5406

TTCTTCGATT GAAACTTCCC ACATGTGCAG GTGCGTTCGC TGTCACTGAT AGACCAAGAG 5466

ACTGAAAGCT TTCACAAATT GCCCTCAAAT CTTCTGTTTC TATCGTCATG ACTCCATATC 5526

TCCGACCACT GGTCATGAGC CAGAGCCCAC TGATTTTGAG GGAATTGGGC TAACCATTTC 5586

CGAGCTTCTG AGTCCTTCTT TTTGATGTCC TTTATGTAGG AATCAAATTC TTCCTTCTGA 5646

CTTGTGGAT 5655

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 594 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Asp Thr Thr Ile Asp Gly Phe Ala Asp Ser Tyr Glu Ile Ser Ser
1 5 10 15

Thr Ser Phe Val Ala Thr Asp Asn Thr Asp Ser Ser Ile Val Tyr Leu
20 25 30

Ala Ala Glu Gln Val Leu Thr Gly Pro Asp Val Ser Ala Leu Gln Leu
35 40 45

Leu Ser Asn Ser Phe Glu Ser Val Phe Asp Ser Pro Asp Asp Phe Tyr
50 55 60

Ser Asp Ala Lys Leu Val Leu Ser Asp Gly Arg Glu Val Ser Phe His
65 70 75 80

Arg Cys Val Leu Ser Ala Arg Ser Ser Phe Phe Lys Ser Ala Leu Ala
85 90 95

Ala Ala Lys Lys Glu Lys Asp Ser Asn Asn Thr Ala Ala Val Lys Leu
100 105 110

Glu Leu Lys Glu Ile Ala Lys Asp Tyr Glu Val Gly Phe Asp Ser Val
115 120 125

Val Thr Val Leu Ala Tyr Val Tyr Ser Ser Arg Val Arg Pro Pro Pro
130 135 140

Lys Gly Val Ser Glu Cys Ala Asp Glu Asn Cys Cys His Val Ala Cys
145 150 155 160

Arg Pro Ala Val Asp Phe Met Leu Glu Val Leu Tyr Leu Ala Phe Ile
165 170 175

Phe Lys Ile Pro Glu Leu Ile Thr Leu Tyr Gln Arg His Leu Leu Asp
180 185 190

Val Val Asp Lys Val Val Ile Glu Asp Thr Leu Val Ile Leu Lys Leu
195 200 205

Ala Asn Ile Cys Gly Lys Ala Cys Met Lys Leu Leu Asp Arg Cys Lys
210 215 220

Glu Ile Ile Val Lys Ser Asn Val Asp Met Val Ser Leu Glu Lys Ser
225 230 235 240

Leu Pro Glu Glu Leu Val Lys Glu Ile Ile Asp Arg Arg Lys Glu Leu
245 250 255

Gly Leu Glu Val Pro Lys Val Lys Lys His Val Ser Asn Val His Lys
260 265 270

Ala Leu Asp Ser Asp Asp Ile Glu Leu Val Lys Leu Leu Leu Lys Glu
275 280 285

Asp His Thr Asn Leu Asp Asp Ala Cys Ala Leu His Phe Ala Val Ala
290 295 300

Tyr Cys Asn Val Lys Thr Ala Thr Asp Leu Leu Lys Leu Asp Leu Ala
305 310 315 320

Asp Val Asn His Arg Asn Pro Arg Gly Tyr Thr Val Leu His Val Ala
325 330 335

Ala Met Arg Lys Glu Pro Gln Leu Ile Leu Ser Leu Leu Glu Lys Gly
340 345 350

Ala Ser Ala Ser Glu Ala Thr Leu Glu Gly Arg Thr Ala Leu Met Ile
355 360 365

Ala Lys Gln Ala Thr Met Ala Val Glu Cys Asn Asn Ile Pro Glu Gln
370 375 380

Cys Lys His Ser Leu Lys Gly Arg Leu Cys Val Glu Ile Leu Glu Gln
385 390 395 400

Glu Asp Lys Arg Glu Gln Ile Pro Arg Asp Val Pro Pro Ser Phe Ala
405 410 415

Val Ala Ala Asp Glu Leu Lys Met Thr Leu Leu Asp Leu Glu Asn Arg
420 425 430

Val Ala Leu Ala Gln Arg Leu Phe Pro Thr Glu Ala Gln Ala Ala Met
435 440 445

Glu Ile Ala Glu Met Lys Gly Thr Cys Glu Phe Ile Val Thr Ser Leu
450 455 460

Glu Pro Asp Arg Leu Thr Gly Thr Lys Arg Thr Ser Pro Gly Val Lys
465 470 475 480

Ile Ala Pro Phe Arg Ile Leu Glu Glu His Gln Ser Arg Leu Lys Ala
485 490 495

Leu Ser Lys Thr Val Glu Leu Gly Lys Arg Phe Phe Pro Arg Cys Ser

```
          500                 505                 510

Ala Val Leu Asp Gln Ile Met Asn Cys Glu Asp Leu Thr Gln Leu Ala
        515                 520                 525

Cys Gly Glu Asp Asp Thr Ala Glu Lys Arg Leu Gln Lys Lys Gln Arg
    530                 535                 540

Tyr Met Glu Ile Gln Glu Thr Leu Lys Lys Ala Phe Ser Glu Asp Asn
545                 550                 555                 560

Leu Glu Leu Gly Asn Ser Ser Leu Thr Asp Ser Thr Ser Ser Thr Ser
                565                 570                 575

Lys Ser Thr Gly Gly Lys Arg Ser Asn Arg Lys Leu Ser His Arg Arg
            580                 585                 590

Arg  *
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Ile Arg Arg Met Arg Arg Ala Leu Asp Ala Ala Asp Ile Glu Leu Val
1               5                  10                  15

Lys Leu Met Val Met Gly Glu Gly Leu Asp Leu Asp Asp Ala Leu Ala
            20                  25                  30

Val His Tyr Ala Val Gln His Cys Asn
        35                  40
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Pro Thr Gly Lys Thr Ala Leu His Leu Ala Ala Glu Met Val Ser Pro
1 5 10 15

Asp Met Val Ser Val Leu Leu Asp His His Ala Asp Xaa Asn Phe Arg
20 25 30

Thr Xaa Asp Gly Val Thr
35

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Ile Arg Arg Met Arg Arg Ala Leu Asp Ala Ala Asp Ile Glu Leu Val
1 5 10 15

Lys Leu Met Val Met Gly Glu Gly Leu Asp Leu Asp Asp Ala Leu Ala
20 25 30

Val His Tyr Ala Val Gln His Cys Asn
35 40

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Arg Arg Pro Asp Ser Lys Thr Ala Leu His Leu Ala Ala Glu Met Val
1 5 10 15

Ser Pro Asp Met Val Ser Val Leu Leu Asp Gln
20 25

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Ile Arg Arg Met Arg Arg Ala Leu Asp Ala Ala Asp Ile Glu Leu Val

1 5 10 15

Lys Leu Met Val Met Gly Glu Gly Leu Asp Leu Asp Asp Ala Leu Ala
20 25 30

Val His Tyr Ala Val Gln His Cys Asn
35 40

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Arg Arg Pro Asp Ser Lys Thr Ala Leu His Leu Ala Ala Glu Met Val
1 5 10 15

Ser Pro Asp Met Val Ser Val Leu Leu Asp Gln
20 25

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Ile Arg Arg Met Arg Arg Ala Leu Asp Ala Ala Asp Ile Glu Leu Val
1 5 10 15

Lys Leu Met Val Met Gly Glu Gly Leu Asp Leu Asp Asp Ala Leu Ala
20 25 30

Val His Tyr Ala Val Gln His Cys Asn
35 40

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Pro Thr Gly Lys Thr Ala Leu His Leu Ala Ala Glu Met Val Ser Pro
1 5 10 15

Asp Met Val

Applicant's or agent's file reference number	Pr. -21212/A/CGC 1909	International application N	PCT/EP 97 / 01 2 18

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page _____, line _____

B. IDENTIFICATION OF DEPOSIT — Further deposits are identified on an additional sheet ☐

Name of depositary institution
American Type Culture Collection (ATCC)

Address of depositary institution *(including postal code and country)*
12301 Parklawn Drive
Rockville, MD 20852
USA

Date of deposit	Accession Number
25 September 1996 (25.09.96)	97736

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* — This information is continued on an additional sheet ☐

We request the Expert Solution where available

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☒ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer: Y. Marinus-v.d. Nouweland	Authorized officer

Form PCT/RO/134 (July 1992)

American Type Culture Collection

12301 Parklawn Drive • Rockville, MD 20852 USA • Telephone: (301)231-5520 Telex: 908-768 ATCCROVE • FAX: 301-816-4366

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF
THE DEPOSIT OF MICROORGANISMS FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL FORM

RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT ISSUED PURSUANT TO RULE 7.3
AND VIABILITY STATEMENT ISSUED PURSUANT TO RULE 10.2

To: (Name and Address of Depositor or Attorney)

Ciba-Geigy Corporation
Attention: Leslie B. Friedrich
P.O. Box 12257
Research Triangle Park, NC 27709

Deposited on Behalf of: Ciba-Geigy Corporation

Identification Reference by Depositor:	ATCC Designation
Cosmid, D7	97736

The deposit was accompanied by: ___ a scientific description ___ a proposed taxonomic description indicated above.

The deposit was received September 25, 1996 by this International Depository Authority and has been accepted.

AT YOUR REQUEST:

X We will inform you of requests for the strain for 30 years.

The strain will be made available if a patent office signatory to the Budapest Treaty certifies one's right to receive, or if a U.S. Patent is issued citing the strain, and ATCC is instructed by the United States Patent & Trademark Office or the depositor to release said strain.

If the culture should die or be destroyed during the effective term of the deposit, it shall be your responsibility to replace it with living culture of the same.

The strain will be maintained for a period of at least 30 years after the date of deposit, and for a period of at least five years after the most recent request for a sample. The United States and many other countries are signatory to the Budapest Treaty.

The viability of the culture cited above was tested October 3, 1996. On that date, the culture was viable.

International Depository Authority: American Type Culture Collection, Rockville, Md. 20852 USA

Signature of person having authority to represent ATCC:
Barbara M. Hailey

Barbara M. Hailey, Administrator, ATCC Patent Depository

Date: October 7, 1996

cc: Andrea C. Walsh, Ph.D.,

Applicant's or agent's file reference number PF, 21212/A/CGC 1909	International application PCT/EP97/01218

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page _____, line _____.

B. IDENTIFICATION OF DEPOSIT — Further deposits are identified on an additional sheet ☐

Name of depositary institution
American Type Culture Collection (ATCC)

Address of depositary institution *(including postal code and country)*
12301 Parklawn Drive
Rockville, MD 20852
USA

Date of deposit	Accession Number
13 June 1996 (13.06.96)	97606

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* — This information is continued on an additional sheet ☐

We request the Expert Solution where available

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☒ This sheet was received with the international application Y. Marinus-v.d. Nouweland	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

Form PCT/RO/134 (July 1992)

American Type Culture Collection

12301 Parklawn Drive • Rockville, MD 20852 USA • Telephone (301)231-5510 Telex: 898-055 ATCCNORTH • FAX: 301-770-2567

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF THE DEPOSIT OF MICROORGANISMS FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL FORM

RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT ISSUED PURSUANT TO RULE 7.3 AND VIABILITY STATEMENT ISSUED PURSUANT TO RULE 10.2

RECEIVED

JUN 27 1996

CIBA-GEIGY
ABRU PATENT DEPT

To: (Name and Address of Depositor or Attorney)

Ciba-Geigy Corporation
Attn: Leslie B. Friedrich
P.O. Box 12257
Research Triangle Park, NC 27709

Deposited on Behalf of: Ciba-Geigy Corporation

Identification Reference by Depositor:	ATCC Designation
Plasmid P1-18	97606

The deposit was accompanied by: ___ a scientific description _ a proposed taxonomic description indicated above.

The deposit was received June 13, 1996 by this International Depository Authority and has been accepted.

AT YOUR REQUEST: X We will inform you of requests for the strain for 30 years.

The strain will be made available if a patent office signatory to the Budapest Treaty certifies one's right to receive, or if a U.S. Patent is issued citing the strain, and ATCC is instructed by the United States Patent & Trademark Office or the depositor to release said strain.

If the culture should die or be destroyed during the effective term of the deposit, it shall be your responsibility to replace it with living culture of the same.

The strain will be maintained for a period of at least 30 years from date of deposit, or five years after the most recent request for a sample, whichever is longer. The United States and many other countries are signatory to the Budapest Treaty.

The viability of the culture cited above was tested June 20, 1996. On that date, the culture was viable.

International Depository Authority: American Type Culture Collection, Rockville, Md. 20852 USA

Signature of person having authority to represent ATCC:

Barbara M. Hailey

Barbara M. Hailey, Administrator, Patent Depository

Date: June 21, 1996

cc: Andrea C. Walsh, Ph.D.

Applicant's or agent's file reference number	P -21212/A/CGC 1909	International application No PCT/EP 97/01218

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page _____, line _____

B. IDENTIFICATION OF DEPOSIT — Further deposits are identified on an additional sheet ☐

Name of depositary institution

American Type Culture Collection (ATCC)

Address of depositary institution (*including postal code and country*)

12301 Parklawn Drive
Rockville, MD 20852
USA

Date of deposit	Accession Number
08 May 1996 (08.05.96)	97543

C. ADDITIONAL INDICATIONS (*leave blank if not applicable*) — This information is continued on an additional sheet ☐

We request the Expert Solution where available

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (*if the indications are not for all designated States*)

E. SEPARATE FURNISHING OF INDICATIONS (*leave blank if not applicable*)

The indications listed below will be submitted to the International Bureau later (*specify the general nature of the indications e.g., "Accession Number of Deposit"*)

For receiving Office use only	For International Bureau use only
☒ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer: Y. Marinus-v.d. Nouweland	Authorized officer:

Form PCT/RO/134 (July 1992)

American Type Culture Collection

12301 Parklawn Drive • Rockville, MD 20852 USA • Telephone: (301)231-5520 Telex: 898-635 ATCCNORTH • FAX: 301-770-2567

A-185
CIBA-161

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF THE DEPOSIT OF MICROORGANISMS FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL FORM

RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT ISSUED PURSUANT TO RULE 7.3 AND VIABILITY STATEMENT ISSUED PURSUANT TO RULE 10.2

RECEIVED
MAY 28 1996
CIBA-GEIGY
ABRU PATENT DEPT.

To: (Name and Address of Depositor or Attorney)

Ciba-Geigy Corporation
Attn: Leslie B. Friedrich
P.O. Box 12267
Research Triangle Park, NC 27709

Deposited on Behalf of: Ciba-Geigy Corporation

Identification Reference by Depositor:	ATCC Designation
Plasmid, BAC4	97543

The deposit was accompanied by: ___ a scientific description _ a proposed taxonomic description indicated above.

The deposit was received May 8, 1996 by this International Depository Authority and has been accepted.

AT YOUR REQUEST: X We will inform you of requests for the strain for 30 years.

The strain will be made available if a patent office signatory to the Budapest Treaty certifies one's right to receive, or if a U.S. Patent is issued citing the strain, and ATCC is instructed by the United States Patent & Trademark Office or the depositor to release said strain.

If the culture should die or be destroyed during the effective term of the deposit, it shall be your responsibility to replace it with living culture of the same.

The strain will be maintained for a period of at least 30 years from date of deposit, or five years after the most recent request for a sample, whichever is longer. The United States and many other countries are signatory to the Budapest Treaty.

The viability of the culture cited above was tested May 17, 1996. On that date, the culture was viable.

International Depository Authority: American Type Culture Collection, Rockville, Md. 20852 USA

Signature of person having authority to represent ATCC:
Barbara M. Hailey

Barbara M. Hailey, Administrator, Patent Depository

Date: May 20, 1996

cc: Andrea C. Walsh, Ph.D.

What Is Claimed Is:

1. An isolated DNA molecule comprising a *NIM1* gene.

2. An isolated DNA molecule according to claim 1, comprising the nucleotide sequence set forth in SEQ ID NO:2.

3. An isolated DNA molecule of about 9.9kb which encodes the *NIM1* gene product.

4. An isolated DNA molecule according to claim 1, comprising the nucleotide sequence set forth in SEQ ID NO:1.

5. An isolated DNA molecule of claim 1, encoding the aminoacid sequence of the NIM1 gene product set forth in SEQ ID NO:2.

6. An isolated DNA molecule comprising a mutant gene of *NIM1* of claim 1, which is a nim1 gene.

7. Clone BAC-04, ATCC Deposit No. 97543.

8. A chimeric gene comprising a promotor active in plant operably linked to a heterologuous DNA molecule encoding the aminoacid sequence of a *NIM1* gene product.

9. A chimeric gene comprising a promotor active in plant operably linked to the heterologuous DNA fragment according to claim 3.

10. A chimeric gene comprising a promotor active in plant operably linked to a heterologuous DNA molecule encoding the aminoacid sequence set forth in SEQ ID NO:2.

11. A chimeric gene comprising a promotor active in plant operably linked to a heterologuous DNA molecule encoding the aminoacid sequence of a nim1 gene product.

12. A recombinant vector comprising the chimeric gene of anyone of claims 8 to 11.

13. A recombinant vector according to claim 12, wherein said vector is capable of being stably transformed into a host cell.

14. A recombinant vector according to claim 12, wherein said vector is capable of being stably transformed into a plant, plant seeds, plant tissue or plant cell.

26. Use of an isolated DNA molecule according to claim 1 and variants thereof in a screening method for identifying compounds capable of inducing broad spectrum disease resistance in plants.

27. Use of a plant phenotyp according to claim 17 to identify an isolated gene fragment which allows expression of broad spectrum of disease resistance in plants.

28. Use of an isolated DNA molecule according to anyone of claims 1 to 5 to confer disease resistance to plant cells, plants and the progeny thereof.

29. Use of an isolated DNA molecule according to claim 6 to confer universal disease susceptibility to plant cells, plants and the progeny thereof.

30. Use of resistant plants and the progeny thereof according to claim 20 to incorporate the resistant trait into plant lines through breeding.

15. A plant expression cassette comprising a chimeric gene of anyone of claims 8 to 11.

16. A plant expression cassette comprising a chimeric gene of claim 8 to 10.

17. A plant expression cassette comprising a chimeric gene of claim 11.

18. A plant expression cassette according to claim 15 to 17 which expresses the chimeric gene continously or constitutively.

19. A plant, plant cells and the progeny thereof comprising the chimeric gene of anyone of claims 8 to 11.

20. A plant, plant cells and the progeny thereof comprising the chimeric gene of anyone of claims 8 to 10, which have a broad spectrum of disease resistance.

21. A plant, plant cells and the progeny thereof comprising the chimeric gene of claim 11.

22. A plant, plant cells and the progeny thereof of claim 19, wherein said plant is selected form the group consisting of gymnosperms, monocots, and dicots.

23. A plant, plant cells and the progeny thereof of claim 19, wherein said plant is a crop plant.

24. A plant, plant cells and the progeny thereof of claim 23, wherein said plant is selected form the group consisting of rice, wheat, barley, rye, corn, potato, carrot, sweet potato, sugar beet, bean, pea, chicory, lettuce, cabbage, cauliflower, broccoli, turnip, radish, spinach, asparagus, onion, garlic, eggplant, pepper, celery, carrot, squash, pumpkin, zucchini, cucumber, apple, pear, quince, melon, plum, cherry, peach, nectarine, apricot, strawberry, grape, raspberry, blackberry, pineapple, avocado, papaya, mango, banana, soybean, tobacco, tomato, sorghum and sugarcane.

25. Use of an isolated DNA molecule according to claim 1, gene to confer disease resistance in plants.

Fig. 1

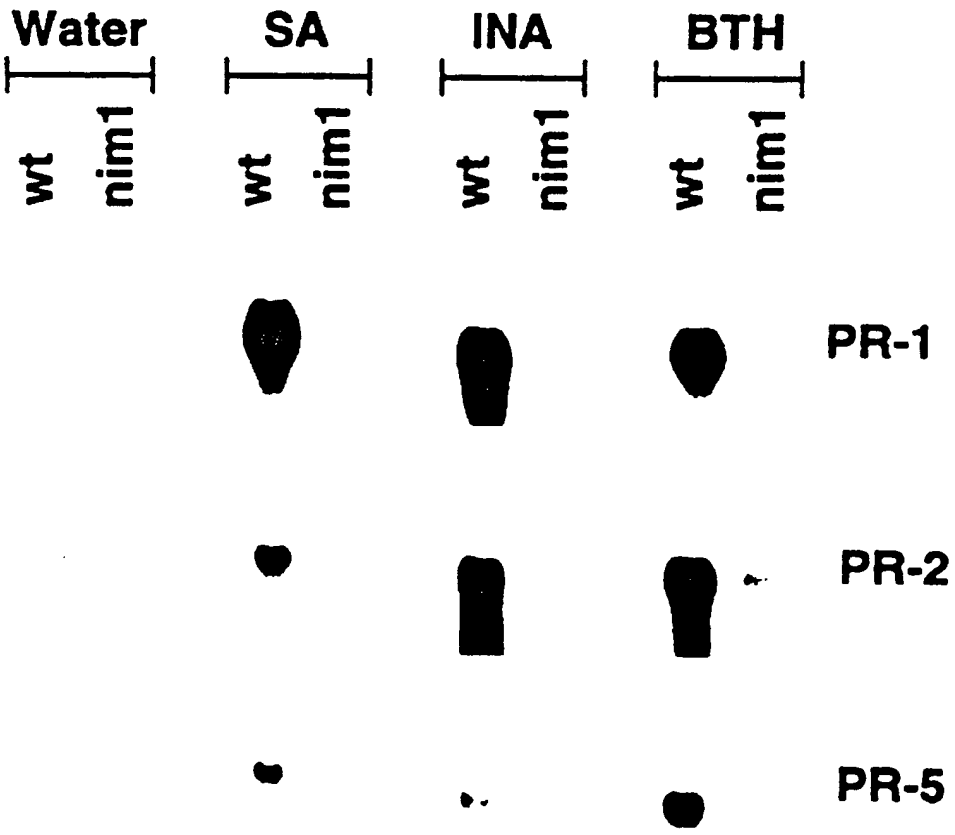


Fig. 2

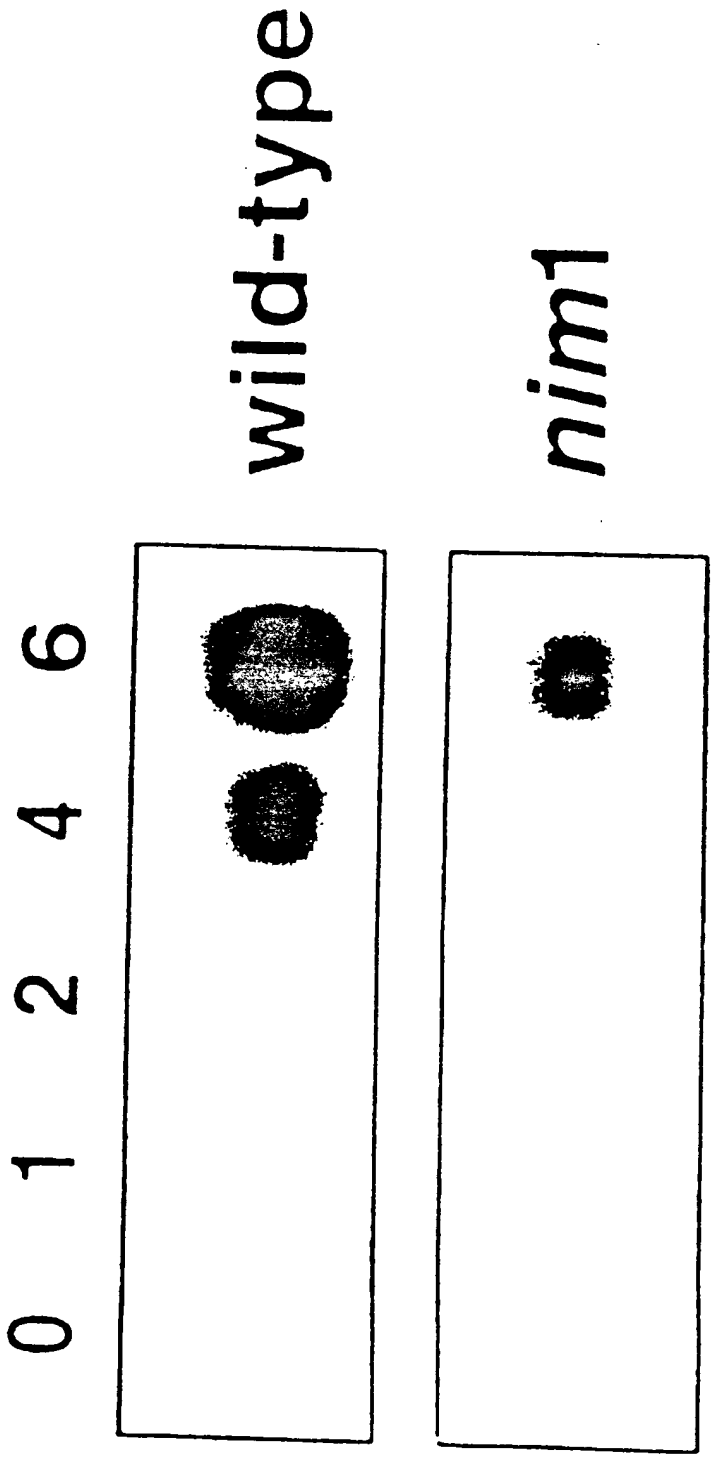
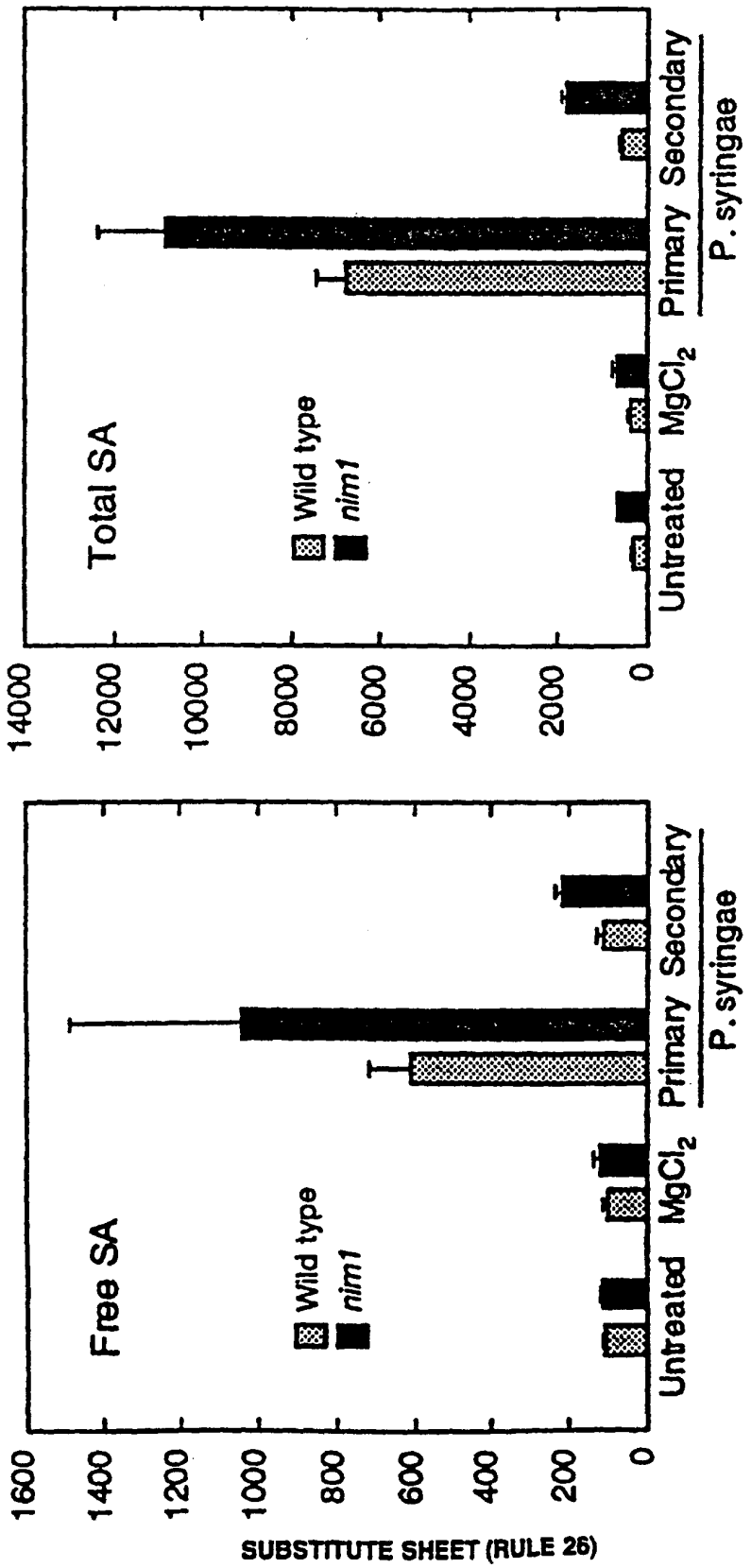


Fig. 3


Total SA
Free SA
Wild type
nim1
Untreated
MgCl2
Primary
Secondary
P. syringae
0
200
400
600
800
1000
1200
1400
1600
2000
4000
6000
8000
10000
12000
14000

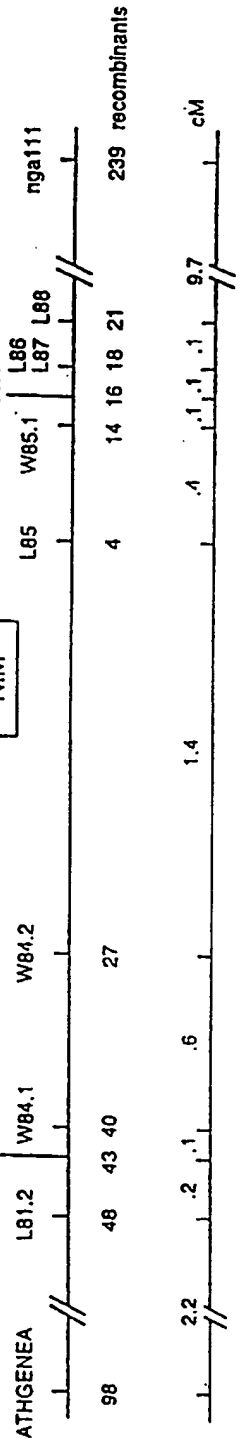
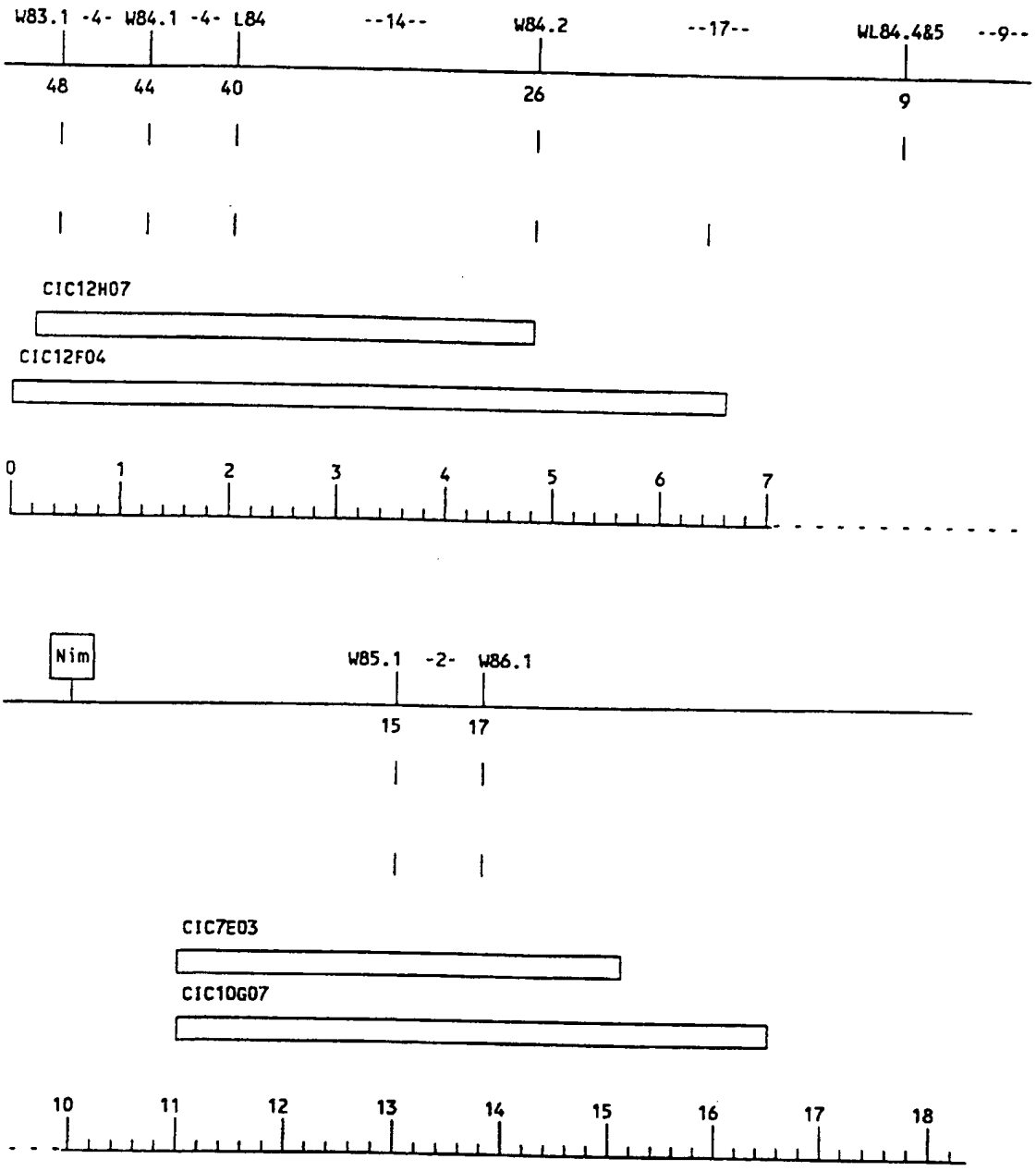


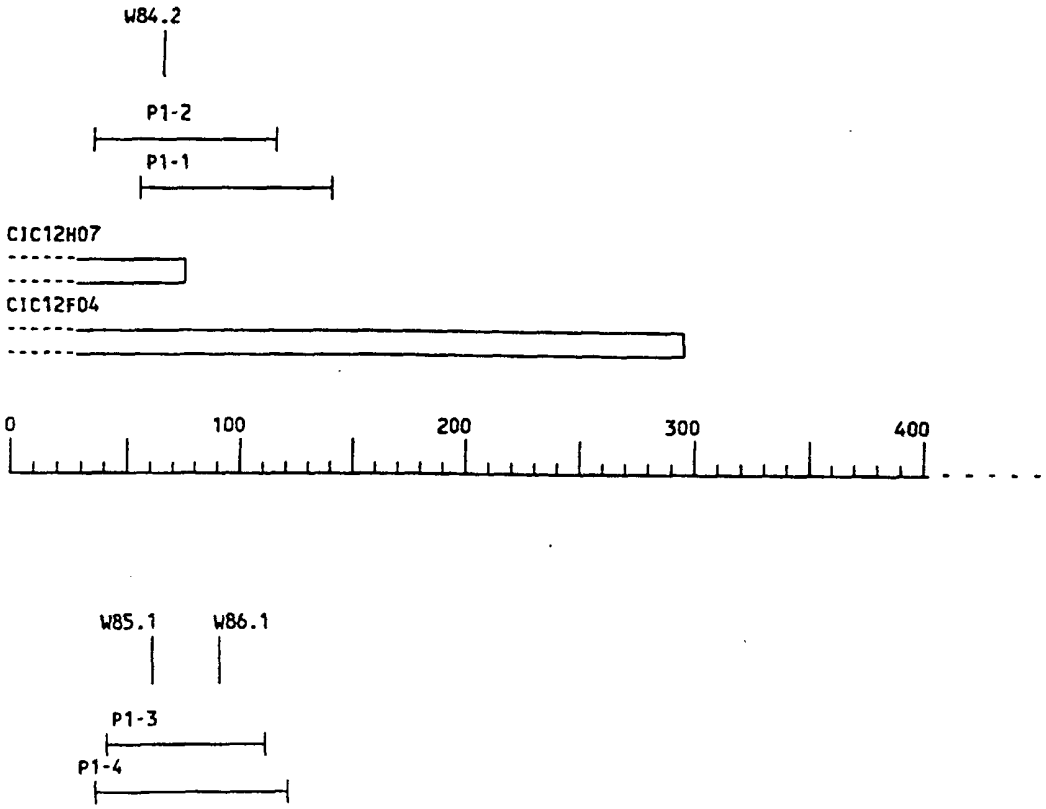
Figure 4. Genetic map of the NIM region, showing the approximate position of NIM, positions of the markers, the number of recombinants that are identified by each marker, and the genetic distances (cM) between the markers.

fig. 5



Schematic representation of the Nim flanking YACs with regard to the genetic map and the AFLP markers. Below the physical distance is indicated in n x 100 kb. Only the positions of markers W84.2 and WL84.4&5 were determined very accurately, the other markers can only be positioned within a certain interval.

Fig. 6



Positions of P1 and BAC clones with respect to the flanking AFLP markers and YACs. The positions of clones P1-3 and P1-4 with respect to YACs 10G07 and 7E03 were not determined. At the bottom the physical distance in kb's is indicated. The physical distance and extent of overlaps are best estimates and not exact values.

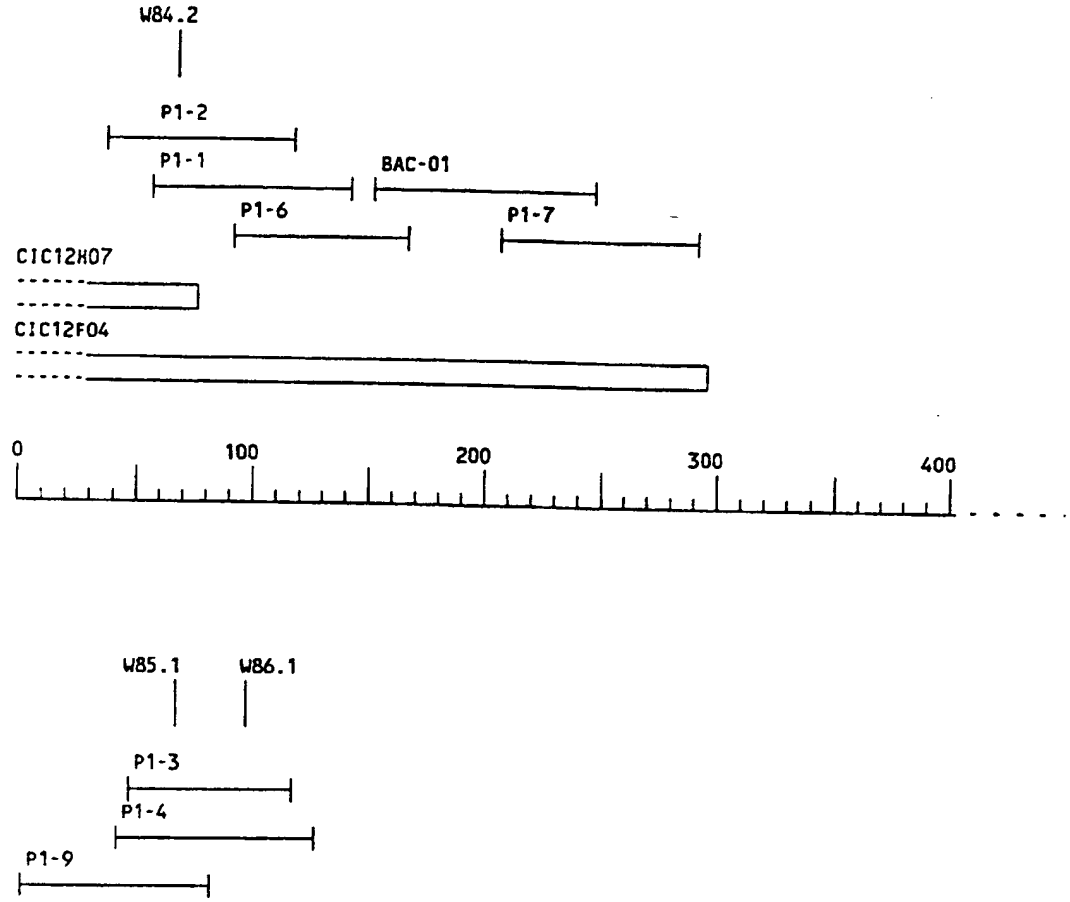


Fig. 7

Extended P1 and BAC contig covering the South end of YAC CIC12F04 and the flanking markers. Clones P1-3, P1-4 and P1-9 overlapped completely with YACs 10G07 and 7E03 and were not positioned with respect to these YACs. At the bottom the physical distance in kb's is indicated. The physical distance and extent of ov rlaps are best estimates and not exact values.

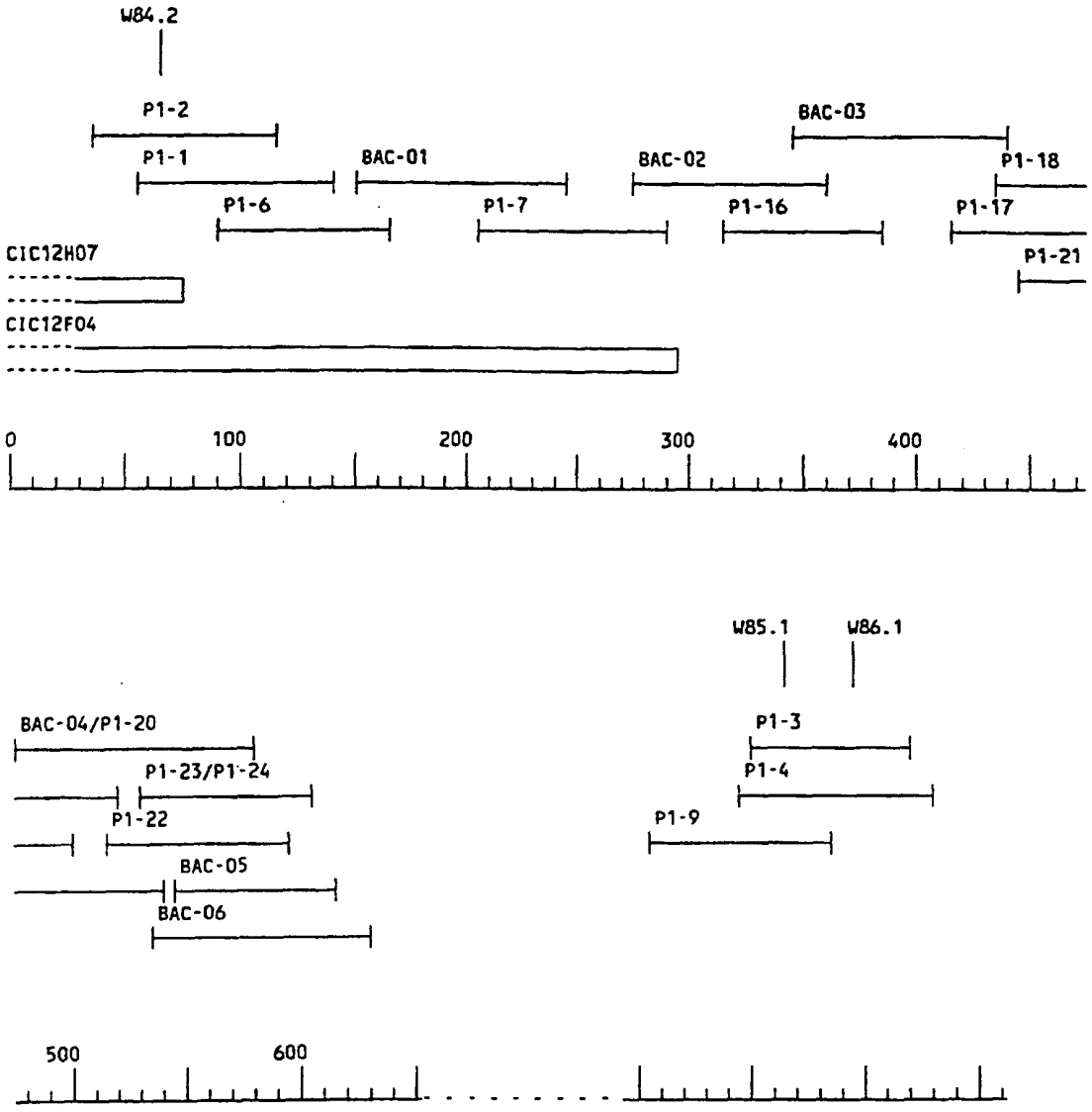


Fig. 8

Schematic representation of all identified P1 and BAC clones and their relative positions in the Nim region. At the bottom the physical distance in kb's is indicated. The physical distance and extent of overlaps are best estimates and not exact values.

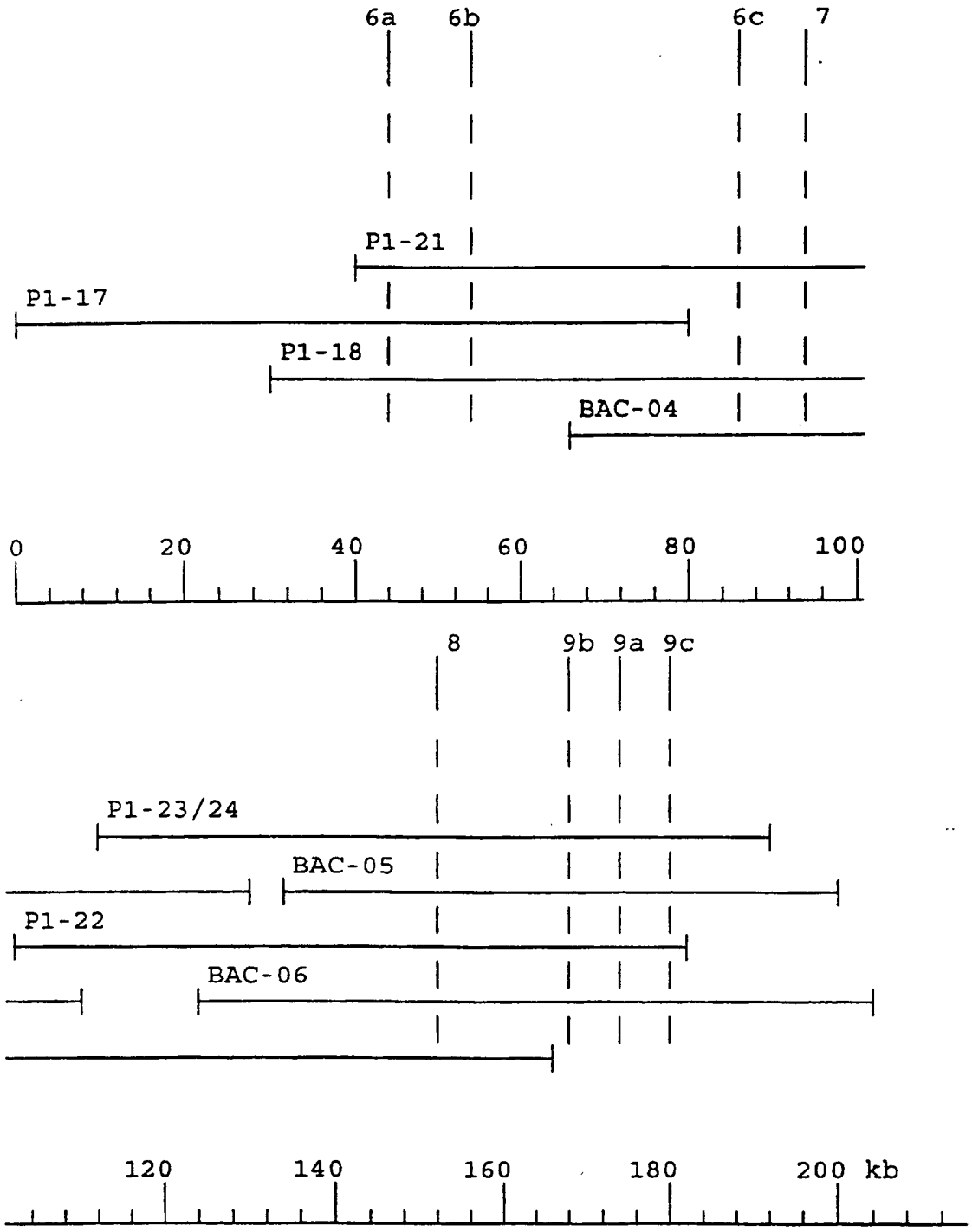


Fig. 9
Integrated genetic and physical fine map of the NIM region. The scale on the bottom is in kb.

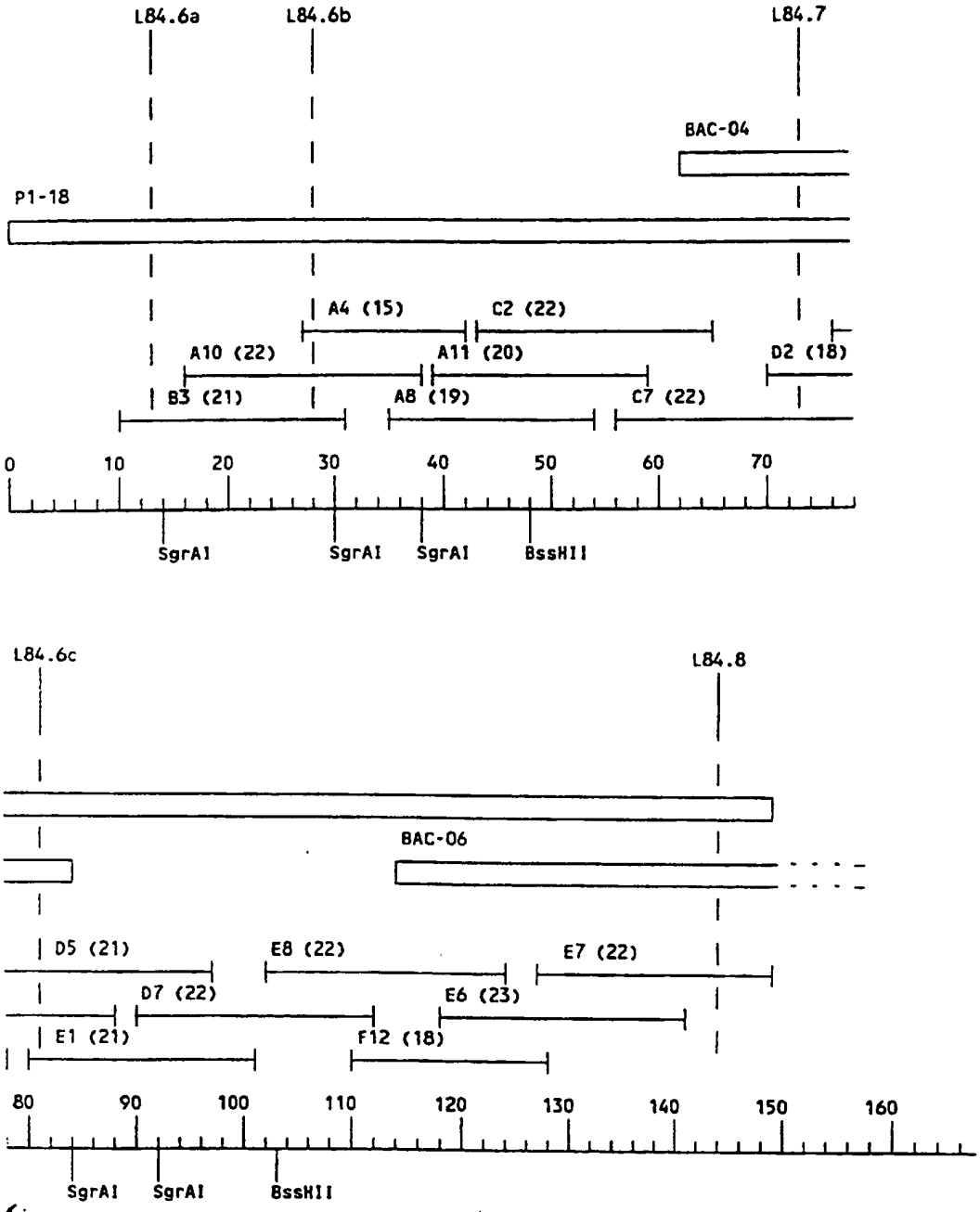


Fig. 10

Integrated map of the NIM region. This map is a best fit of the AFLP fingerprinting and restriction mapping data. Shown are the relevant AFLP markers, the BAC and P1 clones, the cosmid contig, the sizes of the cosmid inserts (in parentheses) and some restriction sites. The scale on th bottom is in kb.

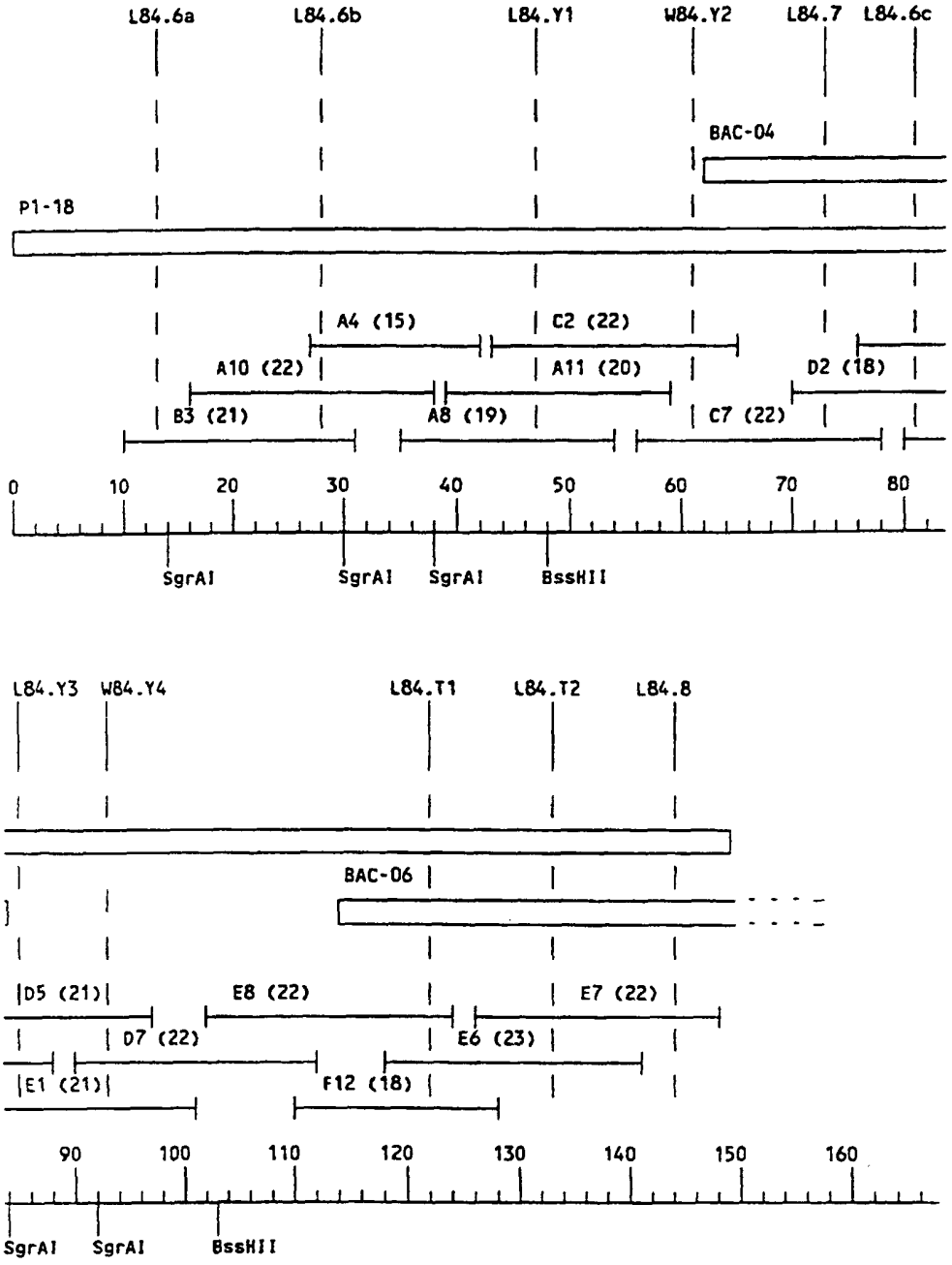


Fig. 11

Integrated map of the NIM region including the new AFLP markers. This map is a best fit of the AFLP fingerprinting and restriction mapping data. Shown are the relevant AFLP markers, the BAC and P1 clones, the cosmid contig, the sizes of the cosmid inserts (in parentheses) and some restriction sites. The scale on the bottom is in kb.

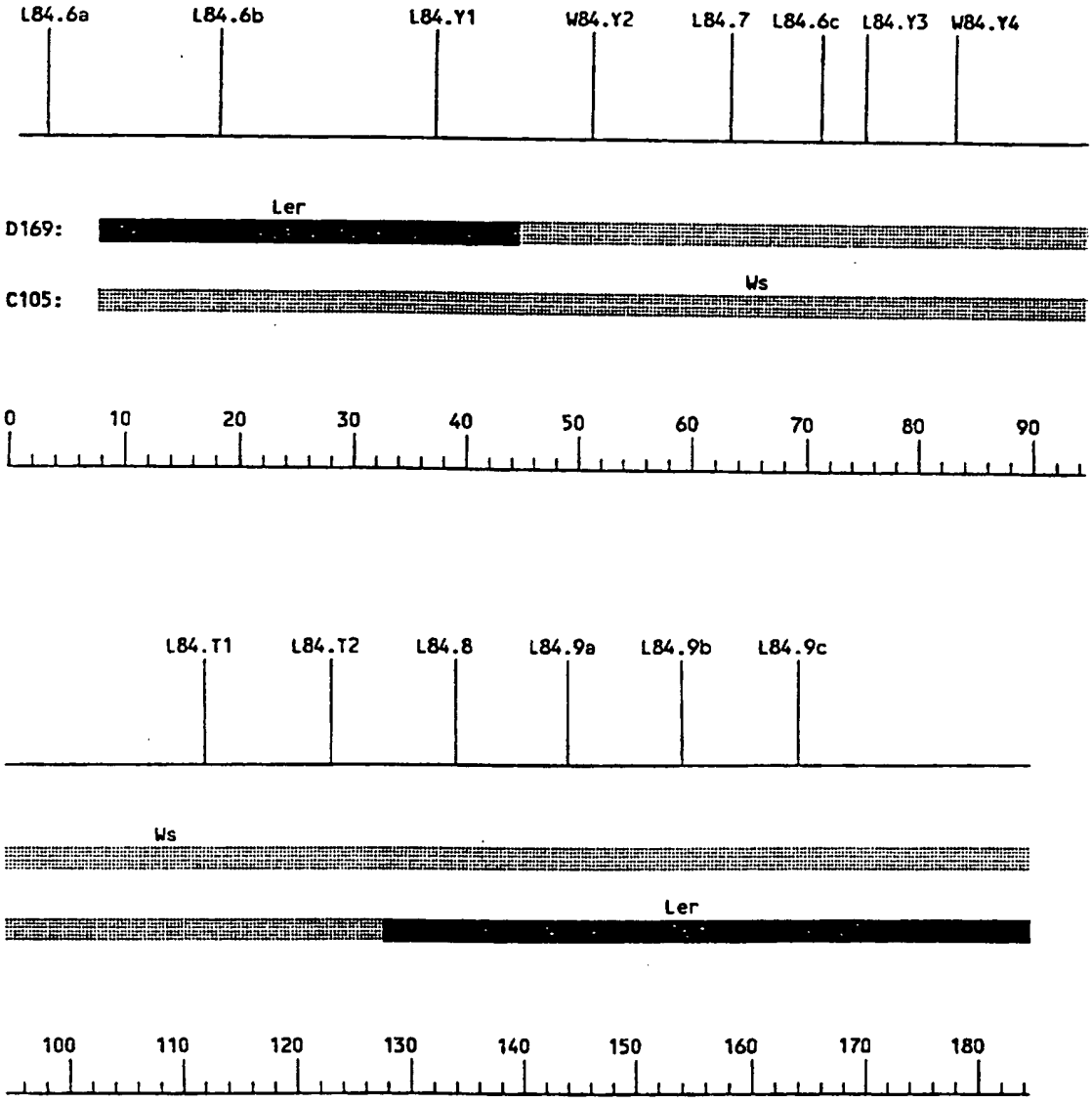
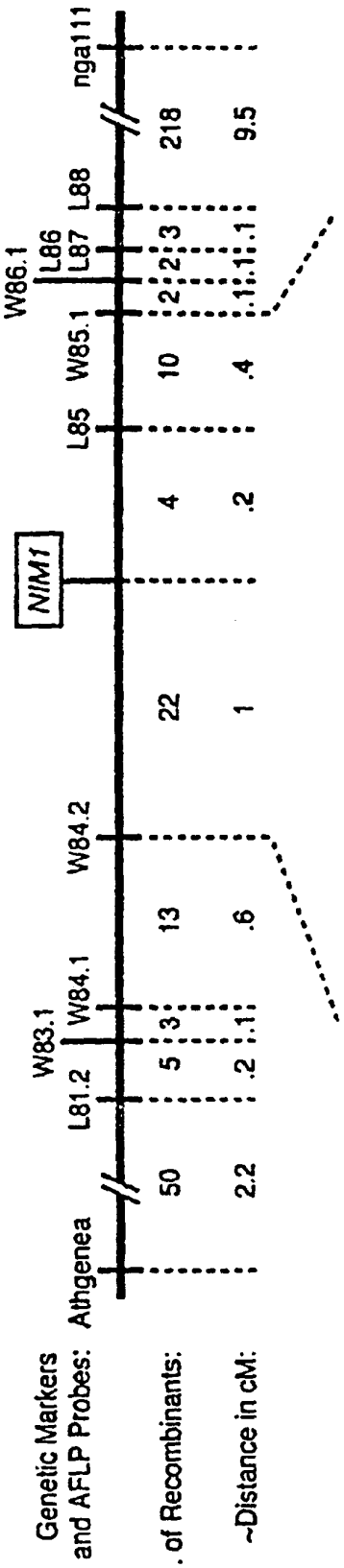
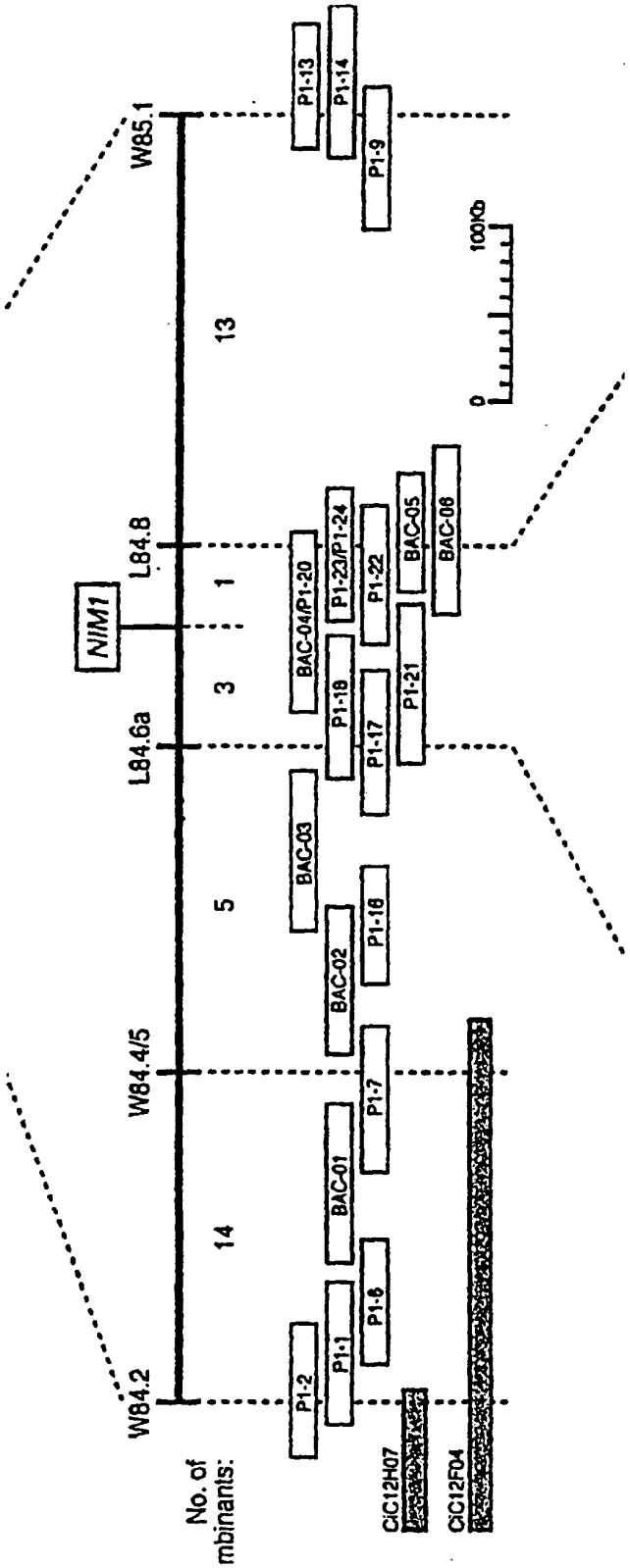


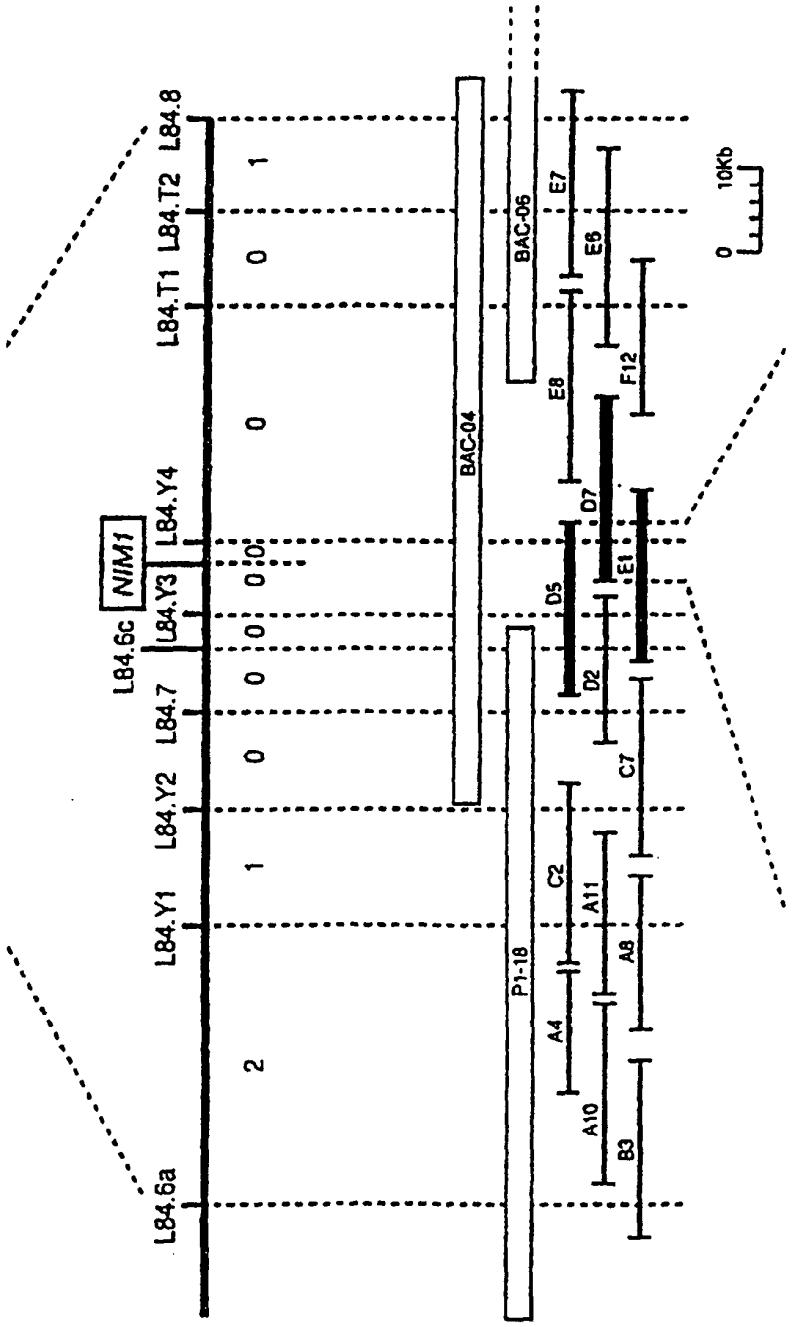
Fig. 12

Schematic representation of the genotypes of the most important recombinants D169 and C105. Shown is the recombinant chromosome with the Ler portion indicated as a black bar and the Ws portion as a gray bar (the non-recombinant chromosome is of the Ws-type). On top the positions and approximate distances of the relevant AFLP markers are shown related to the actual size in kb's as shown at the bottom.

Fig. 13


Genetic Markers and AFLP Probes:
Athgenea
L81.2
W83.1
W84.1
W84.2
NIM1
L85
W85.1
W86.1
L86
L87
L88
nga111
# of Recombinants:
50
5
3
13
22
4
10
2
2
3
218
~Distance in cM:
2.2
.2
.1
.6
1
.2
.4
.1
.1
.1
9.5


W85.1
L84.8
NIM1
L84.6a
W84.4/5
W84.2
No. of recombinants:
13
1
3
5
14
P1-13
P1-14
P1-9
100Kb
0
BAC-04/P1-20
P1-23/P1-24
P1-22
BAC-05
BAC-06
P1-18
P1-17
P1-21
BAC-03
BAC-02
P1-16
P1-7
BAC-01
P1-6
P1-1
P1-2
CIC12H07
CIC12F04



No. of Recombinants:

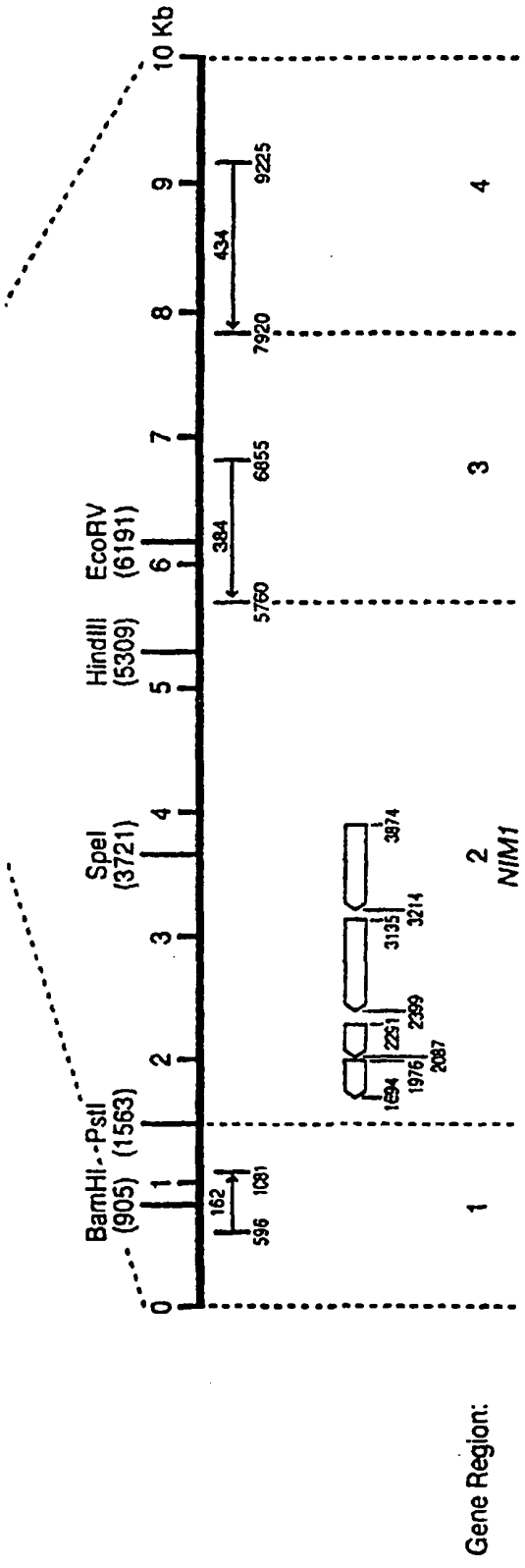

10 Kb
9
8
7
6
5
4
3
2
1
0
BamHI-PstI (905) (1563)
SpeI (3721)
HindIII (5309)
EcoRV (6191)
162
596
1081
384
5760
6855
434
7920
9225
1694
1976
2087
2251
2399
3135
3214
3874
1
2
3
4
NIM1
Gene Region:

Fig. 14

Niml Length: 9919

```
   1  tgatcatgaa ttgcgtgtag ggttgtgttt taaagatagg gatgagctga
  51  agaaggcggt ggactggtgt tccattagag ggcagcaaaa gtgtgtagta
 101  caagagattg agaaggacga gtatacgttt aaatgcatca gatggaaatg
 151  caattggtcg cgtcgggcag attgaataga agaacatgga cttgttaaga
 201  taactaagtg tagttggtcc acatacttgt tgttctatta agccggaaaa
 251  cttcaacttg taatttgcag cagaagagat tgagtgtctg atcagggtac
 301  aacccactct aacagcagag ttgaaaagtt tggtgacatg cttaaaactt
 351  caaagctgcg ggcagcagaa caggaagtaa tcaaagatca gagtttcaga
 401  gtattgccta aactaattgg ctgcatttca ctcatctaat gggctacttg
 451  tggactgcaa tatgagcttt tccctaatcc tgaatttgca tccttcggtg
 501  gcgcgttttg ggcgtttcca cagtccattg aagggtttca acactgtaga
 551  cctctgatca tagtggattc aaaagacttg aacggcaagt accctatgaa
 601  attgatgatt tcctcaggac tcgacgctga tgattgcttt ttcccgcttg
 651  cctttccgct taccaaagaa gtgtccactg atagttggcg ttggtttctc
 701  actaatatca gagagaaggt aacacaaagg aaagacgttt gcctcgtctc
 751  cagtcctcac ccggacatag ttgctgttat taacgaaccc ggatcactgt
 801  ggcaagaacc ttgggtctat cacaggttct gtctggattg tttttgctta
 851  caattccatg atatttttgg agactacaac ctggtgagcc ttgtgaagca
 901  ggctggatcc acaagtcaga aggaagaatt tgattcctac ataaaggaca
 951  tcaaaaagaa ggactcagaa gctcggaaat ggttagccca attccctcaa
1001  aatcagtggg ctctggctca tgaccagtgg tcggagatat ggagtcatga
1051  cgatagaaac agaagatttg agggcaattt gtgaaagctt tcagtctctt
1101  ggtctatcag tgacagcgaa cgcacctgca catgtgggaa gtttcaatcg
1151  aagaagtttc catgtatgca cccagaaatg gtgcaaagga ttgttaactt
1201  gtgtcattca caaatgttgg atgcaatgga gctgactagg agaatgcacc
1251  ttacacgccc actcagtgtt ctcttatctc tagacctgaa actaacttgc
1301  tgtgtaattc gagttacaaa aggttaaagg aagaattagg aagatacata
1351  taacatgaat gttgccagaa gttcagggaa cttgaatatt cttttggttc
1401  ttggtggaaa atatccaaca gatgaacaat ttgacattat ttcacacttt
1451  gattctagca actctgtaac accatcatgg gttattgttg atgtacataa
1501  atatatatta caaatctgta taccattggt tcaaattgtt acaacatttg
1551  tttgaagcac acctgcagca ataatacaca ggatgcaaaa cgaagagcga
1601  aactatatga cgccaacgat agacataaac agttacagtc atcatgaaaa
1651  cagaattata tggtacagca aaaattacac taagaggcaa gagtctcacc
1701  gacgacgatg agagagttta cggttagacc tctttccacc ggttgatttc
1751  gatgtggaag aagtcgaatc tgtcagggac gaatttccta attccaaatt
1801  gtcctcacta aaggccttct ttagtgtctc ttgtatttcc atgtaccttt
1851  gcttcttttg tagtcgtttc tcagcagtgt cgtcttctcc gcaagccagt
```

1901 tgagtcaagt cctcacagtt cataatctgg tcgagcactg ccgaacagcg
1951 cgggaagaat cgtttcccga gttccactga tgataaaaaa aacaaggtca
2001 gacagcaagt aacaaaacca tgtttaaaga tcatttagtt ttgtttttg
2051 tgataaggag tccgatgaag tgggtgagaa tccataccgg ttttagaaag
2101 cgcttttagt ctactttgat gctcttctag gattctgaaa ggtgctatct
2151 ttacacccgg tgatgttctc ttcgtaccag tgagacggtc aggctcgagg
2201 ctagtcacta tgaactcaca tgttcccttc atttcggcga tctccattgc
2251 agcttgtgct tccgttggaa aaagacgttg agcaagtgca actaaacagt
2301 ggacgacaca aagaatagtt atcattagtt cactcagttt cctaatagag
2351 aggacataaa tttaattcaa acatataaga aataagactt gatagatacc
2401 tctattttca agatcgagca gcgtcatctt caattcatcg gccgccactg
2451 caaaagaggg aggaacatct ctaggaattt gttctcgttt gtcttcttgc
2501 tctagtattt ctacacatag tcggcctttg agagaatgct tgcattgctc
2551 cgggatatta ttacattcaa ccgccatagt ggcttgtttt gcgatcatga
2601 gtgcggttct accttccaaa gttgcttctg atgcacttgc acctttttcc
2651 aatagagata gtatcaattg tggctccttc cgcatcgcag caacatgaag
2701 caccgtatat cccctcggat tcctatggtt gacatcggca agatcaagtt
2751 ttaaaagatc tgttgcggtc ttcacattgc aatatgcaac agcgaaatga
2801 agagcacacg catcatctag attggtgtga tcctctttca aaagcaactt
2851 gactaactca atatcatccg agtcaagtgc cttatgtaca ttcgagacat
2901 gtttctttac tttaggtacc tccaaaccaa gctctttacg tctatcaatt
2951 atctctttaa caagctcttc cggcaatgac ttttcaagac taaccatatc
3001 tacattagac ttgacaataa tctctttaca tctatccaat agcttcatac
3051 aagctttacc acatatatta gcaagcttga gtataaccaa tgtgtcctct
3101 ataacaactt tgtctacaac gtccaataag tgcctctgaa atacaaatac
3151 aagtactcaa gtaagaacat attcatgaat gtgtaaccat agcttaatgc
3201 agatggtgtt ttacctgata gagagtaatt aattcaggga tcttgaagat
3251 gaaagccaaa tagagaacct ccaacatgaa atccaccgcc ggccggcaag
3301 ccacgtggca gcaattctcg tctgcgcatt cagaaactcc tttaggcggc
3351 ggtctcactc tgctgctgta aacataagcc aaaacagtca caaccgaatc
3401 gaaaccgact tcgtaatcct tggcaatctc cttaagctcg agcttcacgg
3451 cggcggtgtt gttggagtct ttctccttct tagcggcggc taaagcgctc
3501 ttgaagaaag agcttctcgc tgacaaaacg caccggtgga aagaaacttc
3551 ccggccgtcg gagagaacaa gcttagcgtc gctgtagaaa tcatccggcg
3601 agtcaaagac ggattcgaag ctgttggaga gcaattgcag agcagataca
3651 tcaggtccgg tgagtacttg ttcggcggcc agataaacaa tagaggagtc
3701 ggtgttatcg gtagcgacga aactagtgct gctgatttca taagaatcgg
3751 cgaatccatc aatggtggtg tccatcaaca ggttccgatg aattgaaatt
3801 cacaaattaa agagatctct gctaatcaac gaagagacct tatcaactgg

3851 atttggttaa agatcgaaga taaccattga cgagcagagc caagtcaagt
3901 caacgagagt ggtggtgaga tatgaagaag catcctcgtc ccacggttta
3951 catttcacca aaaccggtaa atttccagga aaggaatctt tgtcagagat
4001 ctttttaaa aagatataac aggaagctaa accggttcgg gttataaatg
4051 ttagtattta taccggagac attttgtgtt gctaattttt gtatatgaga
4101 agttcaatcc ggttcggtaa gcccctgaac caaactagat ttggagatga
4151 tataaatata taaaatttat ttttcatccg gttcgttatt ttcatataaa
4201 tatataaata ttatttttta aatttaagaa ttagatttac atgtgaaagt
4251 tacatttctg tttatttttct ttgaagtaaa atgataaagg gaacgtatat
4301 taagtttcat gctttattca cataagtttt gtaatgtata ttatattttt
4351 cgtttattga aaaagtaatt ttcagtgttc agcatgttta cactataatt
4401 aaatcaagtc gaatatttcc tggaactatt ctccttgttc tatagcaaat
4451 gaaaacgctc ttcacaacaa aatcattata gatataggaa taaattacat
4501 taaaaacatg aaagtcataa tgaatatatt tttttaatta ggatttgatt
4551 taaaaacaat tattgtatac atataaaaga cttctttagt tatttgcctt
4601 caacttctcg ttctgaatca tgcgataaat cagctttttc aataactacg
4651 acgtaaaagc aaattcataa cacgtctaaa caaatttggc tcatccttca
4701 cttgattggt gttttccgga ctcgatgttg ctggaaactg agaagaagaa
4751 ggaatctgca taatcacctc ttggttcctc accggtagac tcattttgtt
4801 ggatcgaaaa cgatcgagat cagaaaatga aaagataggt taaagatgcc
4851 tatgaataca acaacgtaag attatgttga ataaacagag tactttatat
4901 aggagttata ataaggtaaa taaattattg ctttccgcgt tttttacttt
4951 tgtatttctt aaatgataag ttaaattagg ataagatttg tatgatttta
5001 agtaaattta caataactct ctataactca atagcatcac atatttaatt
5051 aattttacta attatctttt gaacaatttt atgaaatagt tttcttttaa
5101 ttaatttttt aaaatgatat attataaaat ttaattgaat caatctgata
5151 taatttttttt atcttctacc atctattata gttgataaat attgtgataa
5201 actttagata aacacccaat tgccaaatat ttaataaatt ttgtgtacca
5251 tgcgtttttt ttggagaata tatatacgtg gacagcatac cgtacatata
5301 ttgtataaaa gcttataaaa catagatacg ggttatattg gtaagctata
5351 aatatatgta aacaatagta agatattacg tgttgtgtct aaatatgtgt
5401 tgctttagat attatgtata tctaatatat taaaatatct tttattaact
5451 aatatattat ttaagagaga aaattgggac actatttttct atacagtaac
5501 tgttttcaac tataaacagg aacccttgat ataataaaat aactagccaa
5551 aaaatcagat taaatattca taaaacaatg tttggtatta ttacataaac
5601 ctaagaaaca aaattcaata ttccttttta ccttataaaa aacaattaaa
5651 catcactaga tatatttatg ccccacaatg agcgagccaa ttgagacttg
5701 agacttgaga tccttgtcaa ctacgtttgc atttgtcggc ccatttttt
5751 tatttttttt ttaaagtgtc ggcccgttgc ttcttccgtt cagatcaacc

5801 ctctcgtaat cagaacaaaa cggaaaacaa acgaaagaac aatcagatcc
5851 ctctttttt gcataaacta aattcaactt ctctgcgttt atgttgtaga
5901 ggcaaccacg atcactacta cgaaacaata caacgtcgtt gcttggagtc
5951 cacgtaatca aatctactcc aatgctttta atatctttca ctttaaccca
6001 cgacttttca aaactgctct ttaaaaccca taactcgtga acatcttctt
6051 gatctttgtt tgtccactga cgaatagcac ctagcttccc ttcgtatctg
6101 actaatcctg agaaaacatc agagttcgga gtatggaaga aggaccaagt
6151 ttcggttttg agacaaaacc ggatcacatt gttgttccgt gatatccaat
6201 gcaagaaccc cgaaacttgt atcgggttgg aaaaaattaa tctgtctgtt
6251 tttggtagac gcaaattttc taatctcttc caggtaaacg aatcagaatc
6301 gaaaacttcg cacataaaag ttctgtgatt caaatggtag ataccccgag
6351 acatacacat acgccgagac tgcgaaagcc tttgtatttt ataccggaaa
6401 gggttcaatc cgattaccgc taaacccaat gacatatccc aacccttcac
6451 ttctggcttt ggtatgacct gatactgttt agtggttggt ttgaagacta
6501 tgtatccacg tgatggtttt gtatacttaa cacaaagcaa tatcccatga
6551 cttgcatcac aagcttcgat ctttatcatt ccgggtggca gaaagtcgat
6601 ggagactcca ttgttttgta aatcactcct ctcatggaca aaactggttc
6651 gaagttcgtg tccttttact atgtagtgtt gtatgaagta tcccgaaata
6701 cgattggttc taaggagatt aagattgaca aaccatgact cgtagcttct
6751 cttgttgcac tctttattca ggagcctgaa ttttccgatt tttgacgccg
6801 gaagataaga aagaaattct tggatcatgt cttgatttat caccggagaa
6851 ctcatgatcc tgtcgggaat aaagagatga gcacgatcac tgaatgagaa
6901 atgaaaaaat caggatcggt agagaacaac ttatgatgaa taaagtgttt
6951 atatatcctt tctttgttta aggaaagtat caaaatttgc ctttttcttc
7001 gctagtccta aaacaaacaa attaaccaaa agataaaatc tttcatgatt
7051 aatgttactt gtgatacctt aagccaaaac tttatcttta gacttttaac
7101 caaatctaca gtaatttaat tgctagactt aggaaacaac ttttttttt
7151 acccaacaat ctttggattt taattgtttt tttttctact aatagattaa
7201 caactcatta tataataatg tttctatcat aattgacaat tctttctttt
7251 taataaacat ccagcttgta taataatcca caagtcaatt tcaccatttt
7301 ggccaattta ttttcttata aaaattagca caaaaaagat tatcattgtt
7351 tagcagattt aatttctaat taacttacgt aatttccatt ttccatagat
7401 ttatctttct ttttatttcc ttagttatct tagtactttc ttagtttcct
7451 tagtaatttt aaattttaag ataatatatt gaaattaaaa gaagaaaaaa
7501 aactctagtt atacttttgt taaatgtttc atcacactaa ctaataattt
7551 tttttagtta aattacaata tataaacact gaagaaagtt tttggcccac
7601 actttttttgg gatcaattag tactatagtt aggggaagat tctgatttaa
7651 aggataccaa aaatgactag ttaggacatg aatgaaaact tataatctca
7701 ataacataca tacgtgttac tgaacaatag taacatctta cgtgttttgt

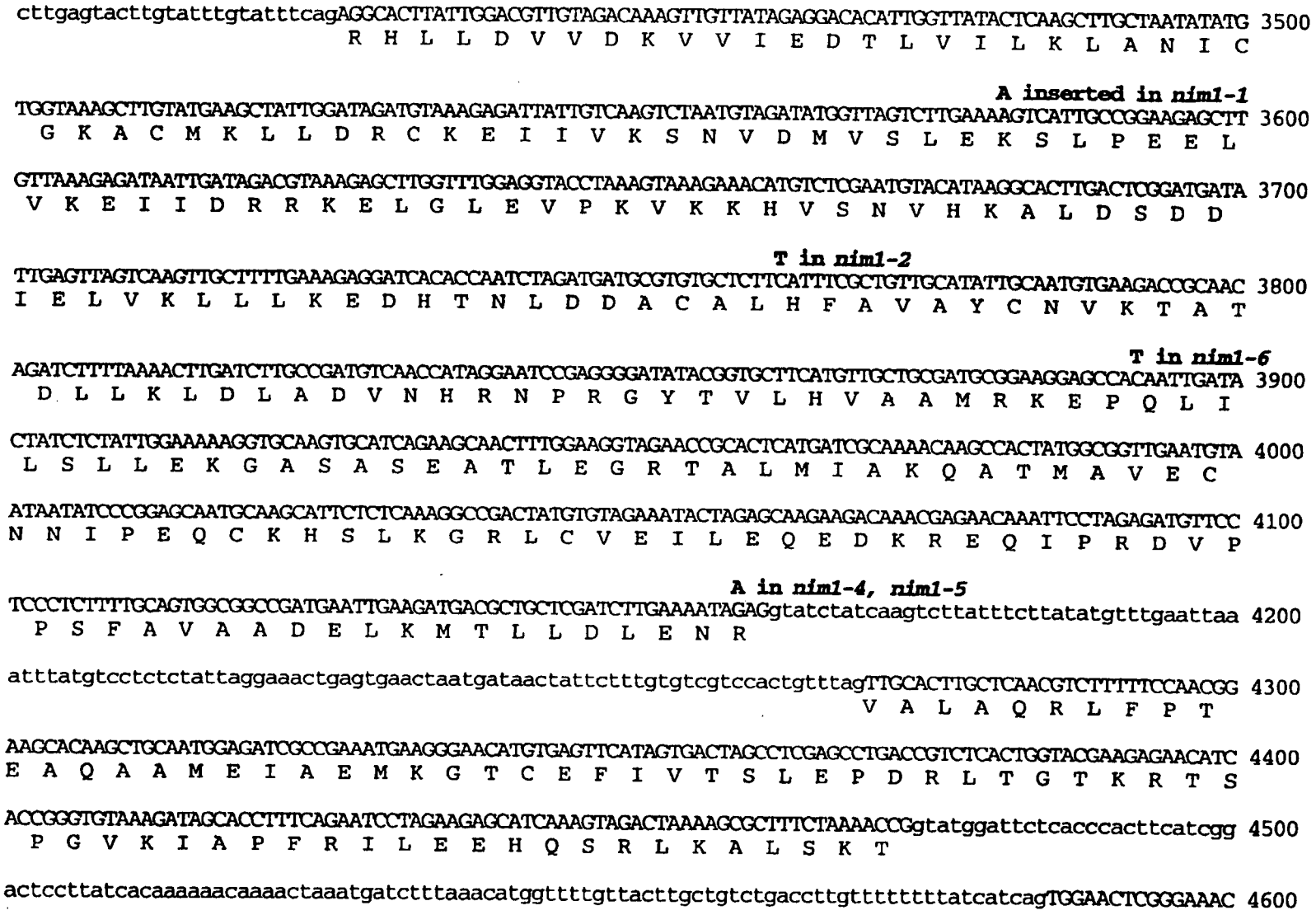
7751 ccatatattt gttgcttata aatatattca tataacaatg tttgcattaa
7801 gcttttaaga agcacaaaac catataacaa aattaaatat tcctatccct
7851 accaaaaaaa aaaattaaat attcctacag ccttgttgat tattttatgc
7901 cctacgttga gccttgttga ctagtttgca tttgtcggtc catttcttct
7951 tccgtccaga tcaaccctct cgtaatcaga acaaaagggg aaacaaacgt
8001 aagaggcaaa atccttgttt gtatgaacta agtttaactt ctctgtgttt
8051 aagttgtaga ggcaaacatg atcccaacta gaaagcatta cgacgtcgtt
8101 gcttggtatc cacgtaatat gctctactcc aatgctttca atatctttca
8151 ctttttccca cgacttttca aaactgctct ttaaaaccca taatctgtga
8201 acatcttctt gattgttgtt tatccagtga cgaataacac ctagcttccc
8251 ttcgtagctg actaactctg ggaataaacc aacgtttgga gtatgtaaga
8301 aagaccaagt ttcggttttg ggacataacc ggatcacatt gtggttccat
8351 gatctccaat gcaagaaccc tgaagcttgt accgggtttg aaagaattag
8401 accgtctgtt ctcggtagac gcaaattttt taatctcttc cacataaacg
8451 aatcggaatc aaaaacttcg cacgcaaaag ttctgagatt ccgagtcata
8501 ccaggcgatt tcgaaagcct aaatatttta taccggaaag gctgcaatcc
8551 ggttaccgtt agacctaatg acttatcaca actcctcact tttgggtttg
8601 gtatgatctg atactgtttt gttgttggtt tgcagactat gtattccggt
8651 attggtcttg tatcattata acaaagcaat atcccatgac gtgcatcaca
8701 agctttgatc tttacctctc cttgtggcag aaaatcgatg gagactcctt
8751 tgttatccaa atctctcctc tcatggaaaa aactggtatc aagtttgtat
8801 cctctttcgt agcgttctag gaagtatcca gagatattgt tggttcgatg
8851 gagatttagg ttgacaaacc aagactcgta gcttctcttg ttgcactctt
8901 tattgatgag cctcaatttt ccgatttcgg acccccgaag ataagaaaga
8951 acctcttgga tcgtgtcctg atttatcacc ggagaactca tgatcttatt
9001 ggaaaaaaga aagaaagaga tgagcacgat cagtgaatga gatatataga
9051 aatcaggatt ggtagagaac cgacgatgat gaatatacaa gtgtttataa
9101 gtatcacaaa ttgccttttt cttcgctagt cccaaaacaa gcaaattaac
9151 caaagataaa atcttcatta atgttttcct ttttcttcgc cagtcccaga
9201 taaaaatata tataaaatat ttcattaggt tacttgtagt accttgagcc
9251 caaagtttct cttttgactt ttaaccaaat taacagtaat ttaatagcta
9301 gacttagaaa acaacatttt gtatatatat tctttgacat caaaattcaa
9351 caatctttgg gtttctatag tgttttttt cttattctaa tagattacca
9401 ctcattatat catatacaaa gtgtttcctt ttcaatcaac atccattttc
9451 tttaaaaatt agcaagtttg ttcttatatc atcattcagc agatttctta
9501 attaaactta gtgatttcca ttttgcacct atatgtttct ctttcttagt
9551 ttagtacttt aaattttcat atatataatt tattaaaatt aaaagtaaaa
9601 actccagttt aacttatgtt aaatgtttca tcacactaaa agagcattaa
9651 gtaataaata ttttagcttt atgaaaaaaa atatcaaatc actgaagaca

```
9701  tttgttggcc tatactctat tttttatttg gccaattagt aatagactaa
9751  tagtaactca tatgatatct ctctaattct ggcgaaacga atattctgat
9801  tctaaagata gtaaaaatga attttgatga agggaatact atttcacaca
9851  cctagaaaga gtaaggtaga aacctttttt tttttggtca gattcttgta
9901  tcaagaagtt ctcatcgat
```

Fig. 15

```
tgtgatgcaagtcatgggatattgctttgtgttaagtatacaaaaccatcacgtggatacatagtcttcaaaccaaccactaaacagtatcaggtcatac 100
caaagccagaagtgaagggttgggatatgtcattgggtttagcggtaatcggattgaacctttccggtataaaatacaaaggctttcgcagtctcggcg 200
tatgtgtatgtctcgggtatcctaccatttgaatcacagaacttttatgtgcgaagttttcgattctgattcgtttacctggaagagattagaaaatttg 300
cgtctaccaaaaacagacagattaatttttttccaacccgatacaagtttcggggttcttgcattggatccacggaacaacaatgtgatccggttttgtc 400
tcaaaaccgaaacttggtcctttcttccatactccgaacttctgatgttttctcaggattagtcagatacgaagggaagctaggtgctattcgtcagtggac 500
aaacaaagatcaagaagatgttcacgagttatgggtttaaagagcagttttgaaaagtcgtgggtttaaagtgaaagatattaaaagcattggagtagat 600
ttgattacgtggactccaagcaacgacgttgtattgtttcgtagtagtgatcgtggttgcctctacaacataaacgcagagaagttgaatttagtttatg 700
caaaaaaagagggatcttgattgttctttcgtttttgtttctgattacgagaggttgatcttgaacggaagaagcaacgggccgacacttta 800
aaaaaaaaataaaaaaaaatgggccgacaaatgcaaacgtagttgacaaggatcttcaagtcttcaattggctcgctcattgtgggcataaatat 900
atctagtgatgtttaattgtttttataaggtaaaaaggaatattgaatttgtttcttaggtttatgtaataataccaaaacattgtttttatgaatattt 1000
aatctgatttttggctagttatttattatatcaagggttcctgtttatagttgaaaacagttactgtatagaaaatagtgtcccaatttttctcttcttta 1100
aataatatattagttaataaaagatattttaatatatctaaagcaacacatatttagacacaacacgtaatatcttactattgt 1200
ttacatatatttatagcttaccaatataacccgtatctatgtttttataagctttttatacaatatatgtacggtatgctgtgtccacgtatatatattctcca 1300
aaaaaaacgcatggtacacaaaatttattaaatatttggcaattgggtgtttatctaaagtttatcacaatatttatcaactatataatagatggtagaaga 1400
taaaaaaattatatcagattgattcaattaaatttttataatatatcatttttaaaaaattaattaaaagaaaactatttcataaaattgttcaaaagataa 1500
ttagtaaaattaattaaatatgtgatgctattgagttatagagagttattgtaaatttacttaaaatcatacaaatctttatcctaatttaacttatatcatt 1600
taagaaatacaaaagtaaaaaacgcggaaagcaataatttatttacctttatttaactcctatatataaagtactctgtttattcaacataatcttacgttg 1700
ttgtattcataggcatctttaacctatctttttcatttttctgatctcgatcgtttttcgatccaacaaaatgagtctaccggtgaggaaccaagaggtgatt 1800
atgcagattccttcttcttcttccagtttccagcaacatcgagtccggaaaacaccaatcaagtgaaggatgagccaaatttgtttagacgtgttatgaatt 1900
```

```
tgcttttacgtcgtagttattgaaaaagctgatttatcgcatgattcagaacgagaagttgaaggcaaataactaaagaagtcttttatatgtatacaat 2000

aattgtttttaaatcaaatcctaattaaaaaaatatattcattatgactttcatgtttttaatgtaatttattcctatatctataatgattttgttgtga 2100

agagcgttttcatttgctatagaacaaggagaatagttccaggaaatattcgacttgatttaattatagtgtaaacatgctgaacactgaaaattacttt 2200

ttcaataaacgaaaaatataatatacattacaaaacttatgtgaataaagcatgaaacttaatatacgttccctttatcattttacttcaaagaaaataa 2300

acagaaatgtaactttcacatgtaaatctaattcttaaatttaaaaaataatatttatatatttatatgaaaataacgaaccggatgaaaaataaatttt 2400

atatatttatatcatctccaaatctagtttggttcaggggcttaccgaaccggattgaacttctcatatacaaaaattagcaacacaaaatgtctccggt 2500

ataaatactaacatttataacccgaaccggtttagcttcctgttatatctttttaaaaaagatctctgacaaagattcctttcctggAAATTTACCGGTT 2600

TTGGTGAAATGTAAACCGTGGGACGAGGATGCTTCTTCATATCTCACCACCACTCTCGTTGACTTGACTTGGCTCTGCTCGTCAATGGTTATCTTCGATC 2700

TTTAACCAAATCCAGTTGATAAGGTCTCTTCGTTGATTAGCAGAGATCTCTTTAATTTGTGAATTTCAATTCATCGGAACCTGTTGATGGACACCACCAT 2800
                                                                                M  D  T  T  I

TGATGGATTCGCCGATTCTTATGAAATCAGCAGCACTAGTTTCGTCGCTACCGATAACACCGACTCCTCTATTGTTTATCTGGCCGCCGAACAAGTACTC 2900
  D  G  F  A  D  S  Y  E  I  S  S  T  S  F  V  A  T  D  N  T  D  S  S  I  V  Y  L  A  A  E  Q  V  L

ACCGGACCTGATGTATCTGCTCTGCAATTGCTCTCCAACAGCTTCGAATCCGTCTTTGACTCGCCGGATGATTTCTACAGCGACGCTAAGCTTGTTCTCT 3000
 T  G  P  D  V  S  A  L  Q  L  L  S  N  S  F  E  S  V  F  D  S  P  D  D  F  Y  S  D  A  K  L  V  L

CCGACGGCCGGGAAGTTTCTTTCCACCGGTGCGTTTTGTCAGCGAGAAGCTCTTTCTTCAAGAGCGCTTTAGCCGCCGCTAAGAAGGAGAAAGACTCCAA 3100
S  D  G  R  E  V  S  F  H  R  C  V  L  S  A  R  S  S  F  F  K  S  A  L  A  A  A  K  K  E  K  D  S  N

CAACACCGCCGCCGTGAAGCTCGAGCTTAAGGAGATTGCCAAGGATTACGAAGTCGGTTTCGATTCGGTTGTGACTGTTTTGGCTTATGTTTACAGCAGC 3200
  N  T  A  A  V  K  L  E  L  K  E  I  A  K  D  Y  E  V  G  F  D  S  V  V  T  V  L  A  Y  V  Y  S  S

AGAGTGAGACCGCCGCCTAAAGGAGTTTCTGAATGCGCAGACGAGAATTGCTGCCACGTGGCTTGCCGGCCGGCGGTGGATTTCATGTTGGAGGTTCTCT 3300
 R  V  R  P  P  P  K  G  V  S  E  C  A  D  E  N  C  C  H  V  A  C  R  P  A  V  D  F  M  L  E  V  L

- deleted in nim1-3
ATTTGGCTTTCATCTTCAAGATCCCTGAATTAATTACTCTCTATCAGgtaaaacaccatctgcattaagctatggttacacattcatgaatatgttctta 3400
Y  L  A  F  I  F  K  I  P  E  L  I  T  L  Y  Q
```


cttgagtacttgtatttgtatttcagAGGCACTTATTGGACGTTGTAGACAAAGTTGTTATAGAGGACACATTGGTTATACTCAAGCTTGCTAATATATG 3500
R H L L D V V D K V V I E D T L V I L K L A N I C
A inserted in nim1-1
TGGTAAAGCTTGTATGAAGCTATTGGATAGATGTAAAGAGATTATTGTCAAGTCTAATGTAGATATGGTTAGTCTTGAAAAGTCATTGCCGGAAGAGCTT 3600
G K A C M K L L D R C K E I I V K S N V D M V S L E K S L P E E L
GTTAAAGAGATAATTGATAGACGTAAAGAGCTTGGTTTGGAGGTACCTAAAGTAAAGAAACATGTCTCGAATGTACATAAGGCACTTGACTCGGATGATA 3700
V K E I I D R R K E L G L E V P K V K K H V S N V H K A L D S D D
T in nim1-2
TTGAGTTAGTCAAGTTGCTTTTGAAAGAGGATCACACCAATCTAGATGATGCGTGTGCTCTTCATTTCGCTGTTGCATATTGCAATGTGAAGACCGCAAC 3800
I E L V K L L L K E D H T N L D D A C A L H F A V A Y C N V K T A T
T in nim1-6
AGATCTTTTAAAACTTGATCTTGCCGATGTCAACCATAGGAATCCGAGGGGATATACGGTGCTTCATGTTGCTGCGATGCGGAAGGAGCCACAATTGATA 3900
D L L K L D L A D V N H R N P R G Y T V L H V A A M R K E P Q L I
CTATCTCTATTGGAAAAAGGTGCAAGTGCATCAGAAGCAACTTTGGAAGGTAGAACCGCACTCATGATCGCAAAACAAGCCACTATGGCGGTTGAATGTA 4000
L S L L E K G A S A S E A T L E G R T A L M I A K Q A T M A V E C
ATAATATCCCGGAGCAATGCAAGCATTCTCTCAAAGGCCGACTATGTGTAGAAATACTAGAGCAAGAAGACAAACGAGAACAAATTCCTAGAGATGTTCC 4100
N N I P E Q C K H S L K G R L C V E I L E Q E D K R E Q I P R D V P
A in nim1-4, nim1-5
TCCCTCTTTTGCAGTGGCGGCCGATGAATTGAAGATGACGCTGCTCGATCTTGAAAATAGAGgtatctatcaagtcttatttcttatatgtttgaattaa 4200
P S F A V A A D E L K M T L L D L E N R
atttatgtcctctctattaggaaactgagtgaactaatgataactattctttgtgtcgtccactgtttagTTGCACTTGCTCAACGTCTTTTTCCAACGG 4300
V A L A Q R L F P T
AAGCACAAGCTGCAATGGAGATCGCCGAAATGAAGGGAACATGTGAGTTCATAGTGACTAGCCTCGAGCCTGACCGTCTCACTGGTACGAAGAGAACATC 4400
E A Q A A M E I A E M K G T C E F I V T S L E P D R L T G T K R T S
ACCGGGTGTAAAGATAGCACCTTTCAGAATCCTAGAAGAGCATCAAAGTAGACTAAAAGCGCTTTCTAAAACCGgtatggattctcacccacttcatcgg 4500
P G V K I A P F R I L E E H Q S R L K A L S K T
actccttatcacaaaaaacaaaactaaatgatctttaaacatggttttgttacttgctgtctgaccttgtttttttttatcatcagTGGAACTCGGGAAAC 4600

V E L G K

GATTCTTCCCCGCGCTGTTCGGCAGTGCTCGACCAGATTATGAACTGTGAGGACTTGACTCAACTGGCTTGCCGAGAAGACGACACTGCTGAGAAACGACT 4700
R F F P R C S A V L D D Q I M N C E D L T Q L A C G E D D T A E K R L

ACAAAGAAGCAAAGGTACATGGAAATACAAGACACTAAAGAAGGCCTTTAGTGAGGACAATTTGGAATTAGGAAATTCGTCCCTGACAGATTCGACT 4800
Q K K Q R Y M E I Q E T L K K A F S E D N L E L G N S S L T D S T

TCTTCCACATCGAAATCAACCGGTGGAAAGAGGTCTAACCGTAAACTCTCTCATCGTCGTCGGTGAGACTCTTGCCTCTTAGTGTAATTTTTGCTGTACC 4900
S S T S K S T G G K R S N R K L S H R R R *

ATATAATTCTGTTTTCATGATGACTGTAACTGTTTATGTCTATCGTTGGCGTCATATAGTTTCGCTCTTCGTTTTGCATCCTGTGTATTATTGCTGCAGG 5000

TGTGCTTCAAACAAATGTTGTAACAATTTGAACCAATGGTATACAGATTTGTAatatatatatttatgtacatcaacaataacccatgatggtgttacagag 5100

ttgctagaatcaaagtgtgaaataatgtcaaattgttcatctgttggatattttccaccaagaaccaaaagaatattcaagttccctgaacttcctggcaa 5200

cattcatgttatatgtatcttcctaattcttcctttaaccttttgtaactcgaattacacagcaagttagtttcaggtctagagataagagaacactgag 5300

tgggcgtgtaaggtgcattctcctagtcagctccattgcatccaacatttgtgaatgacacaagttaacaatccttgcaccatttctgggtgcatacat 5400

ggaaacttcttcgattgaaacttcccacatgtgcagtcgttcgctgtcactgatagaccaagagactgaaagctttcacaaattgccctcaaatcttc 5500

tgtttctatcgtcatgactccatatctccgaccactggtcatgagccagagcccactgattttgagggaattgggctaaccatttccgagcttctgagtc 5600

cttctttttgatgtccttttatgtaggaatcaaattcttccttctgacttgtggat 5655

fig. 17

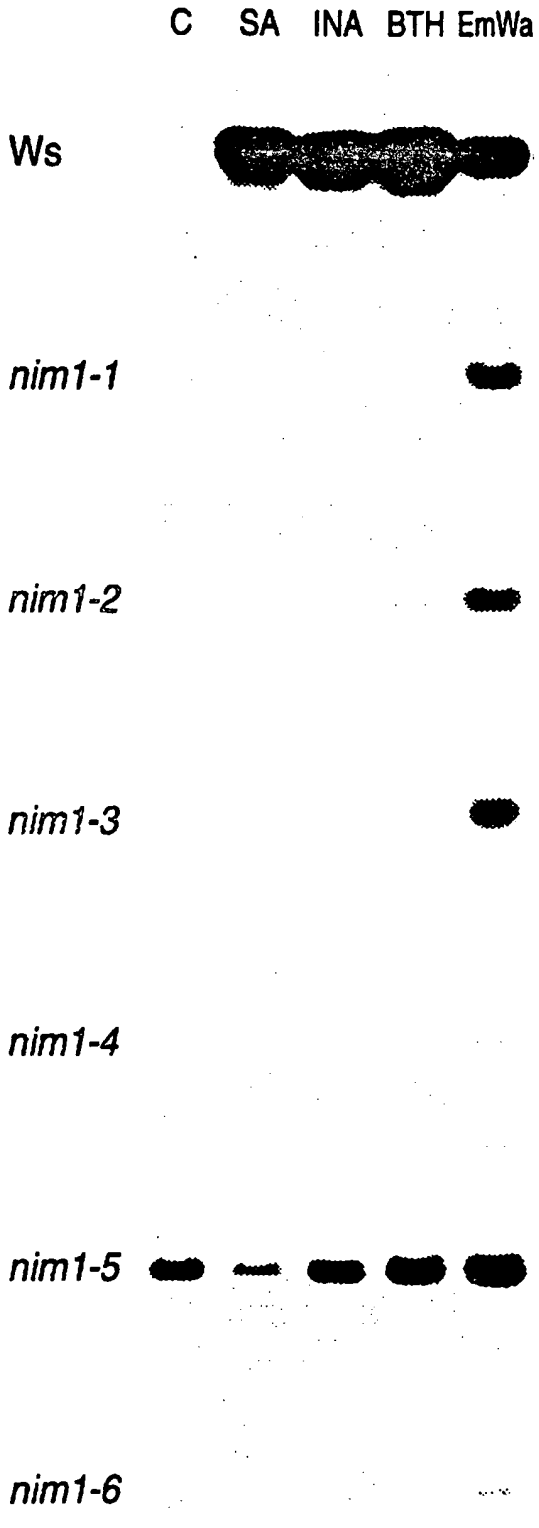


fig. 6

16/19

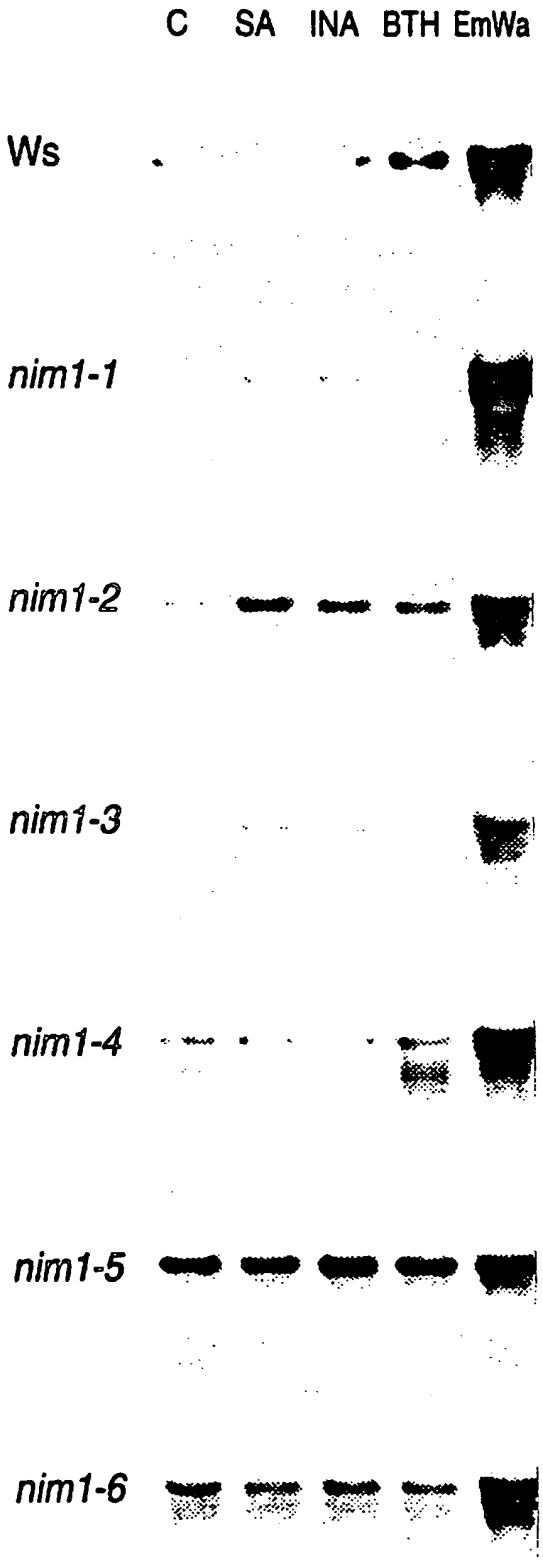
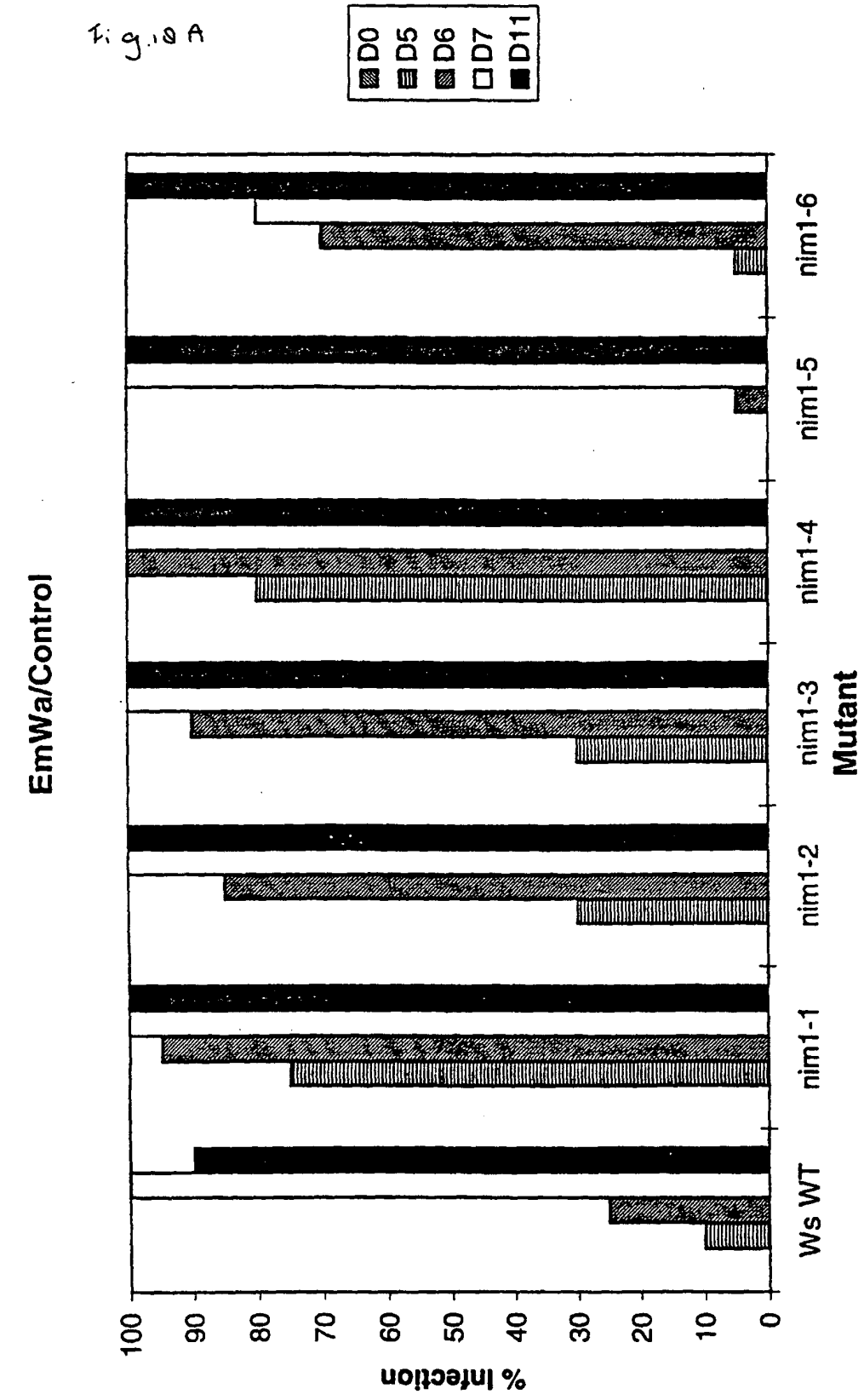
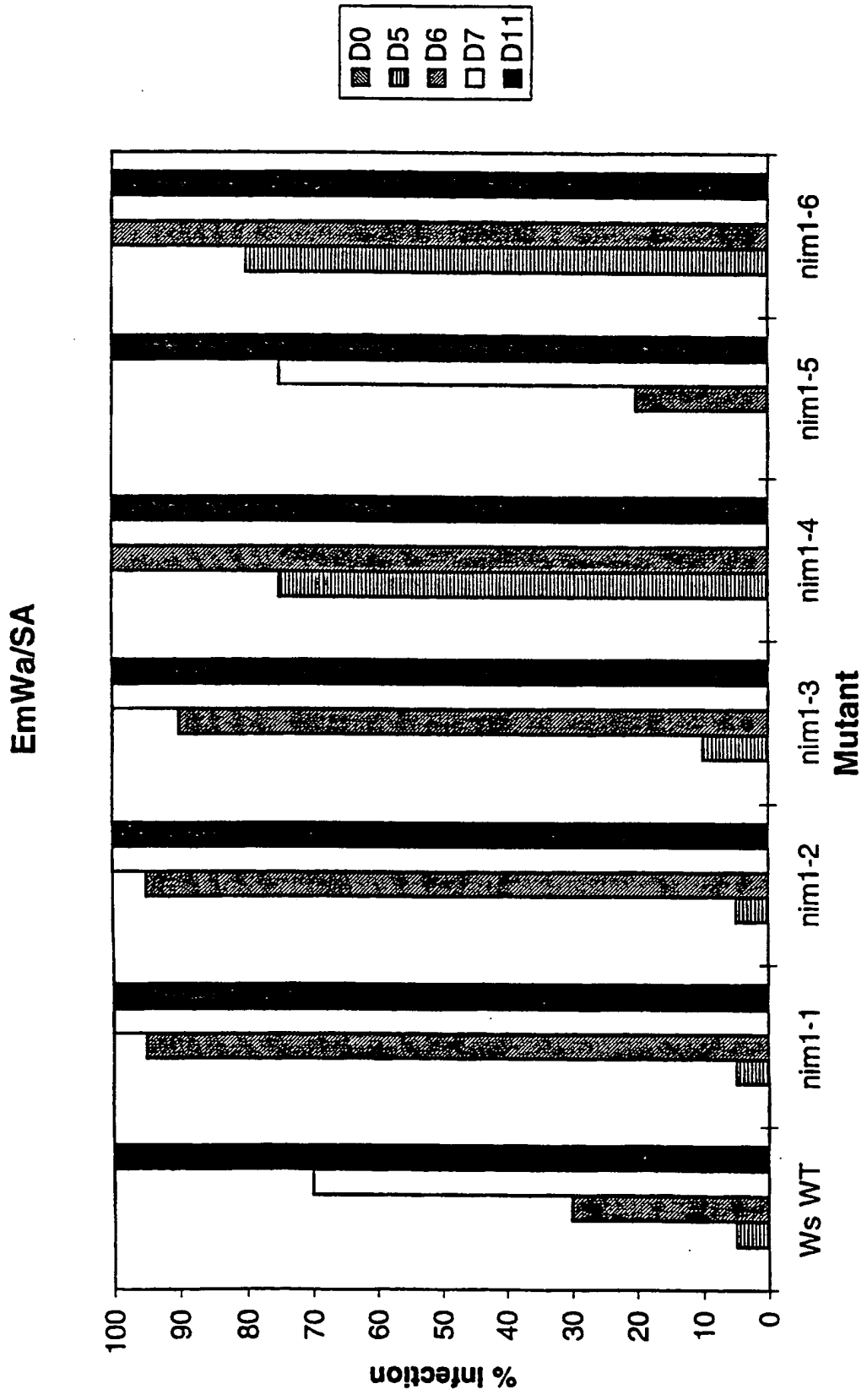
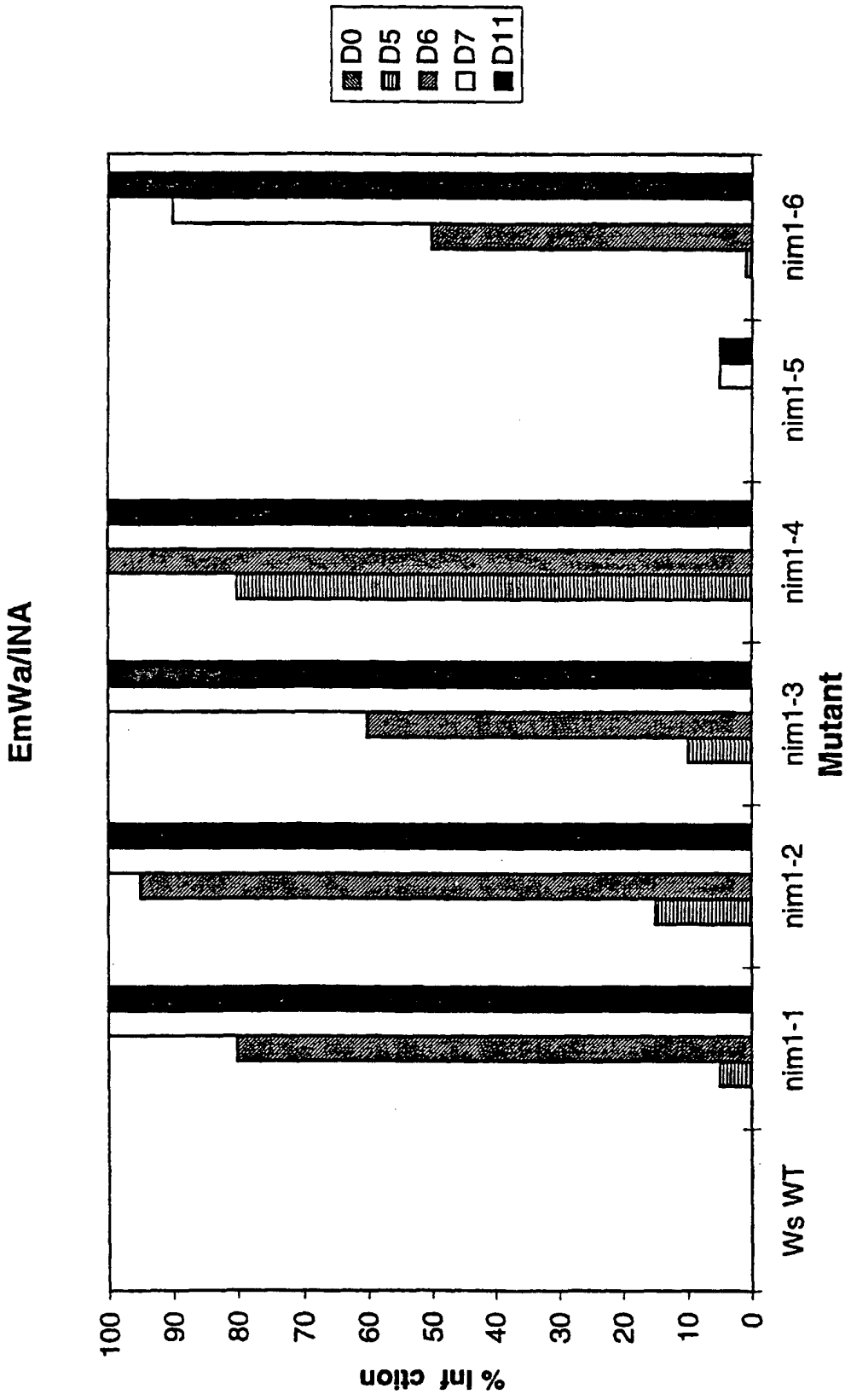


Fig.10A


D0
D5
D6
D7
D11
EmWa/Control
% Infection
100
90
80
70
60
50
40
30
20
10
0
Ws WT
nim1-1
nim1-2
nim1-3
nim1-4
nim1-5
nim1-6
Mutant


D0
D5
D6
D7
D11
EmWa/SA
Ws WT
nim1-1
nim1-2
nim1-3
nim1-4
nim1-5
nim1-6
Mutant
% Infection
100
90
80
70
60
50
40
30
20
10
0


D0
D5
D6
D7
D11
EmWa/INA
Ws WT
nim1-1
nim1-2
nim1-3
nim1-4
nim1-5
nim1-6
Mutant
% Inf ction
100
90
80
70
60
50
40
30
20
10
0

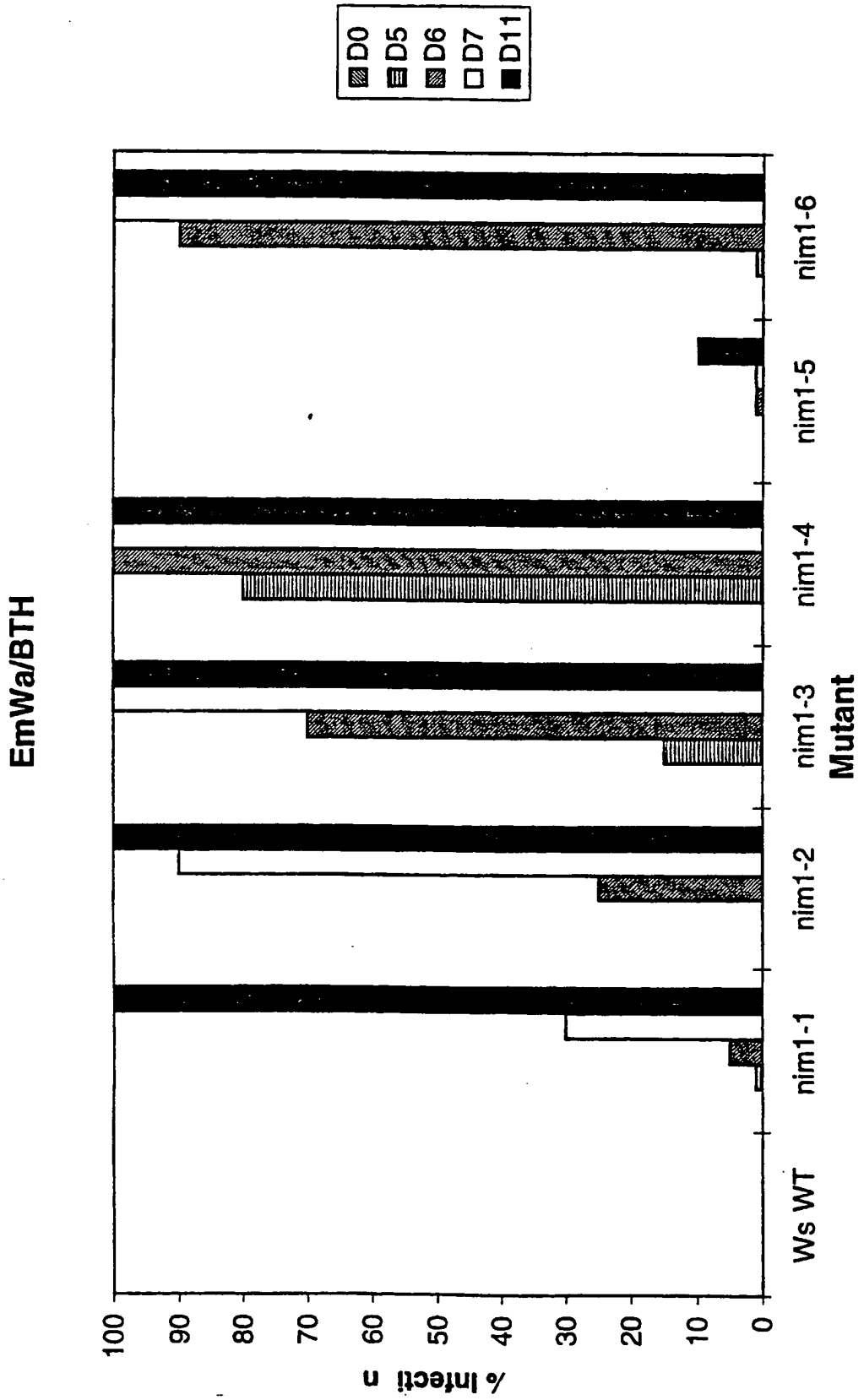
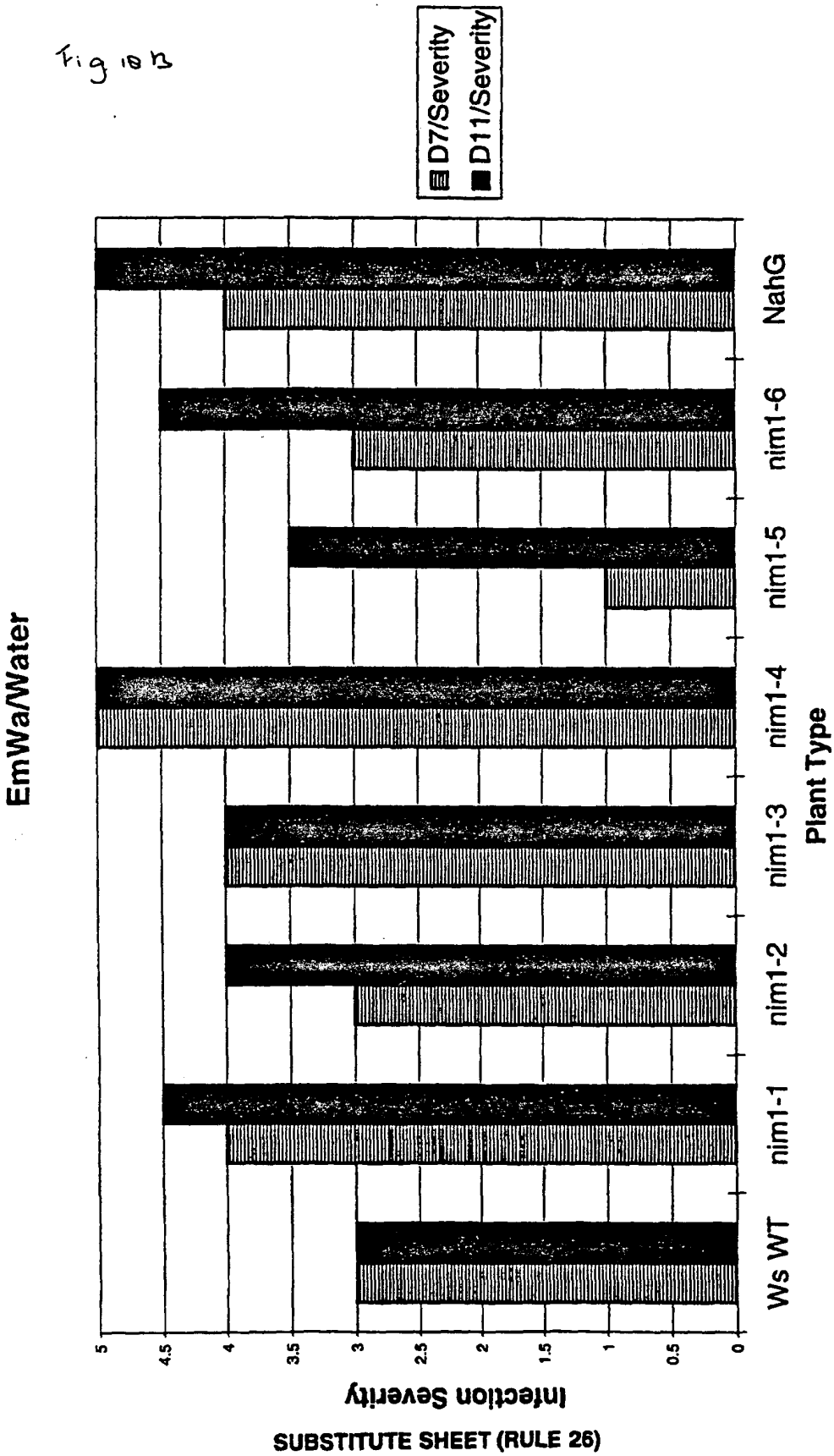
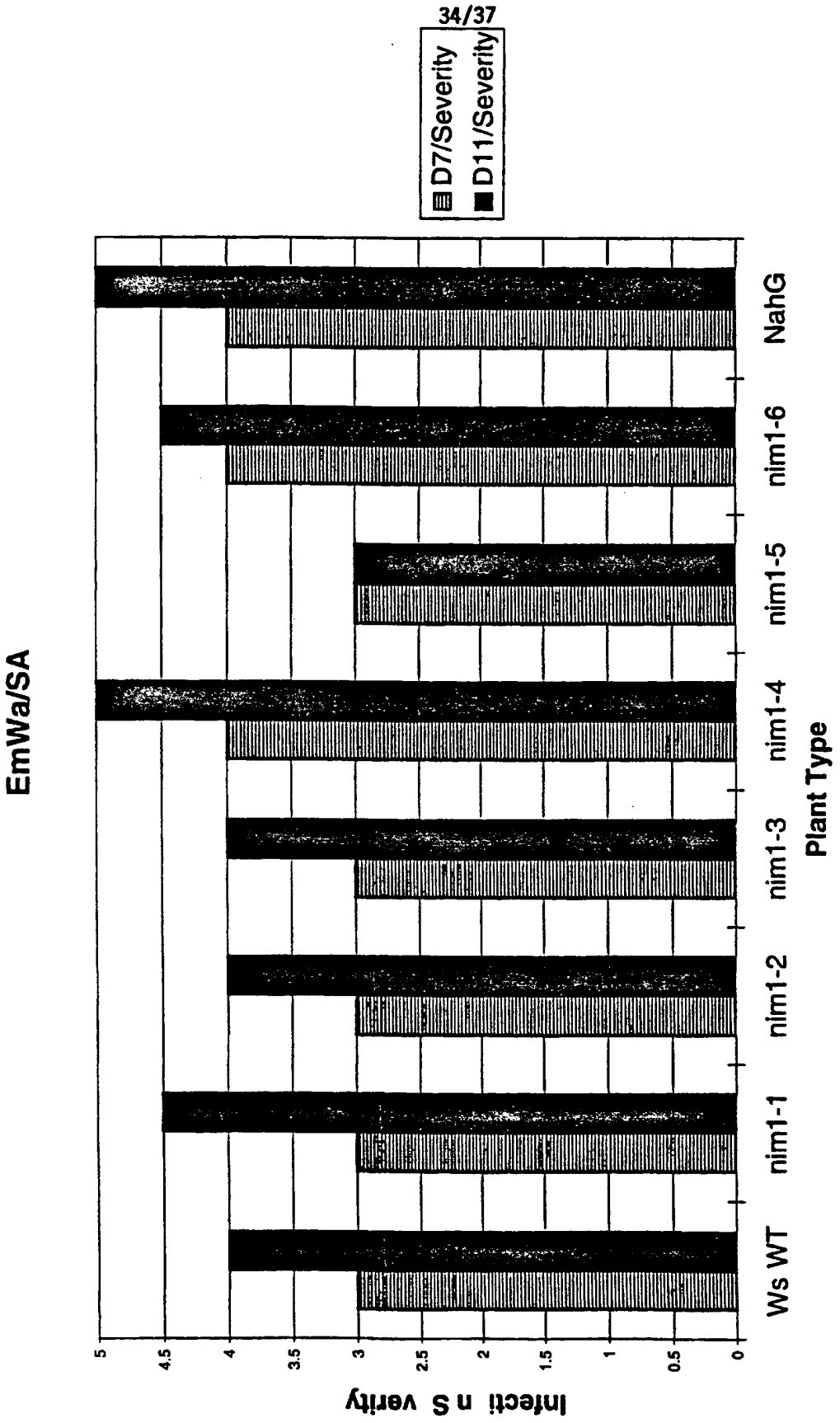
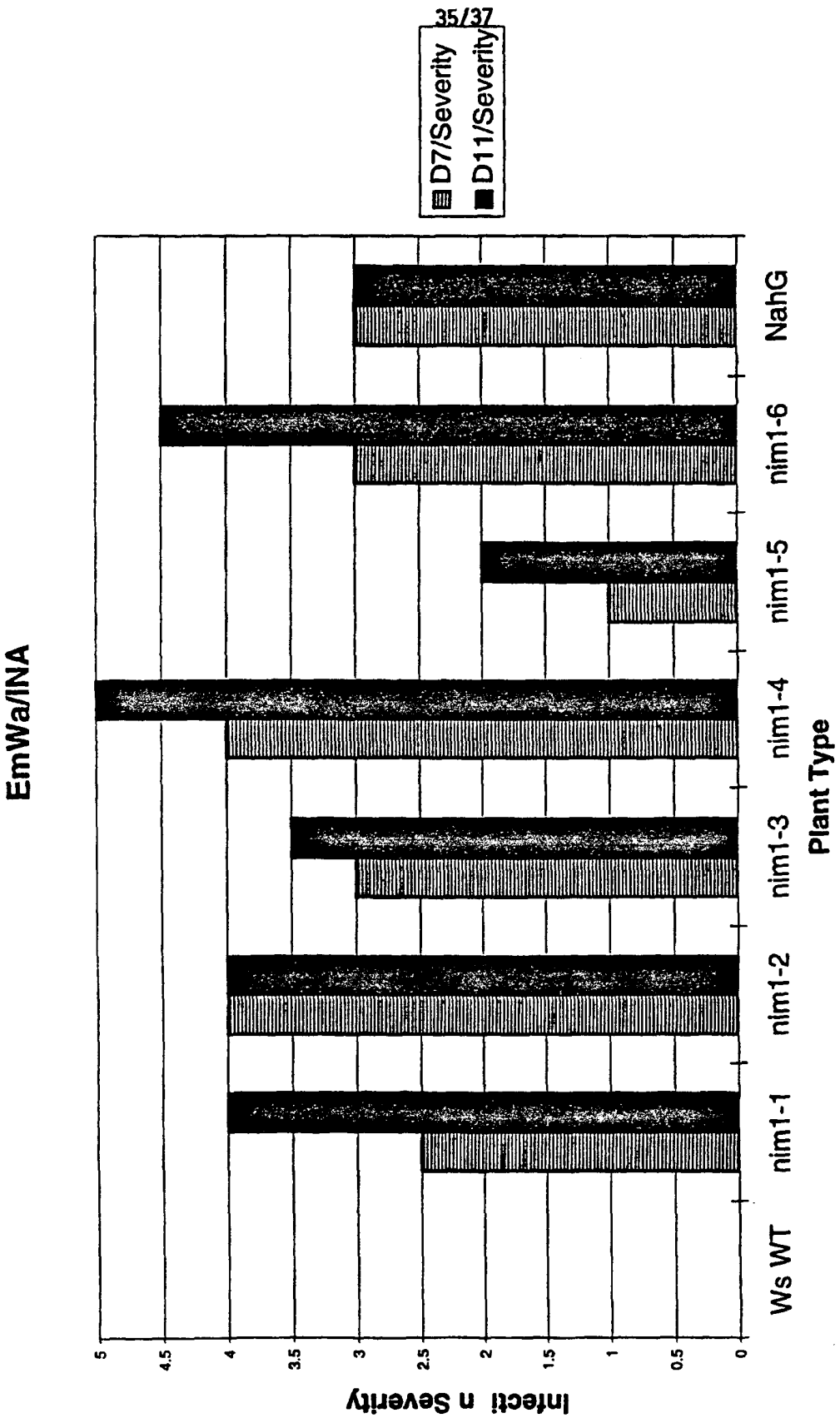
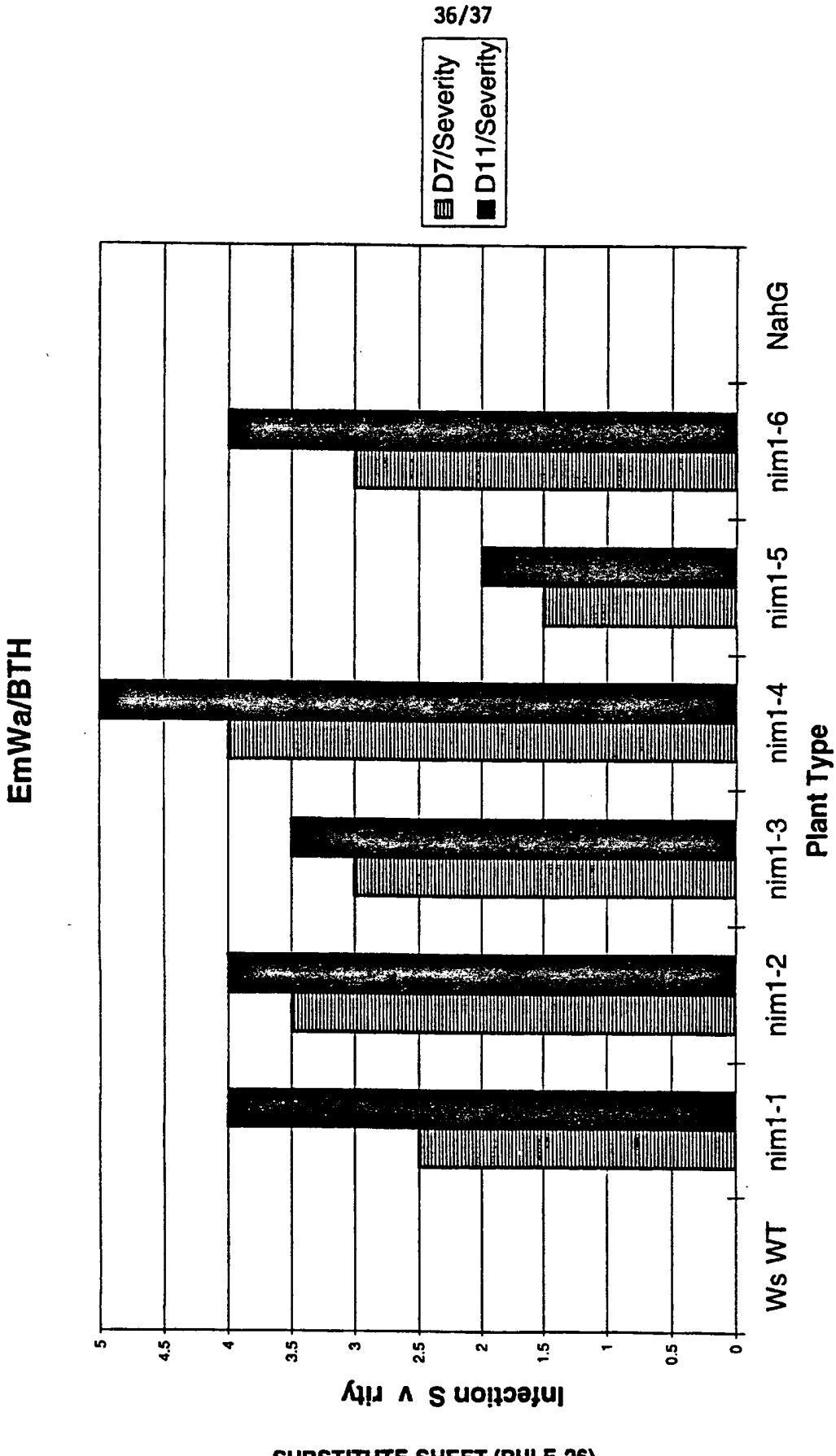

D0
D5
D6
D7
D11
EmWa/BTH
% Infecti n
100
90
80
70
60
50
40
30
20
10
0
Ws WT
nim1-1
nim1-2
nim1-3
nim1-4
nim1-5
nim1-6
Mutant

Fig 10B


D7/Severity
D11/Severity
EmWa/Water
Infection Severity
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Ws WT
nim1-1
nim1-2
nim1-3
nim1-4
nim1-5
nim1-6
NahG
Plant Type


D7/Severity
D11/Severity
EmWa/SA
Infecti n S verity
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Ws WT
nim1-1
nim1-2
nim1-3
nim1-4
nim1-5
nim1-6
NahG
Plant Type


D7/Severity
D11/Severity
EmWa/INA
Infecti n Severity
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Ws WT
nim1-1
nim1-2
nim1-3
nim1-4
nim1-5
nim1-6
NahG
Plant Type


D7/Severity
D11/Severity
EmWa/BTH
Infection S v rity
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Ws WT
nim1-1
nim1-2
nim1-3
nim1-4
nim1-5
nim1-6
NahG
Plant Type

```
nim     :   267 VSNVHKALDSDDIELVKLLLKEDHTNLDDACALHFAVAYCN 307
              +  + +ALD+ DIELVKL++  +  +LDDA A+H+AV +CN
Rice-1  :    33 IRRMRRALDAADIELVKLMVMGEGLDLDDALAVHYAVQHCN 155

nim     :   327 PRGYTVLHVAAMRKEPQLILSLLEKGASASEATLEGRT 364
              P G T LH+AA    P ++  LL+  A  +  T +G T
Rice-1  :   215 PTGKTALHLAAEMVSPDMVSVLLDHHADXNFRTXDGVT 328


nim     :   267 VSNVHKALDSDDIELVKLLLKEDHTNLDDACALHFAVAYCN 307
              +  + +ALD+ DIELVKL++  +  +LDDA A+H+AV +CN
Rice-2  :    33 IRRMRRALDAADIELVKLMVMGEGLDLDDALAVHYAVQHCN 155

nim     :   325 RNPRGYTVLHVAAMRKEPQLILSLLEK 351
              R P   T LH+AA    P ++  LL++
Rice-2  :   208 RRPDSKTALHLAAEMVSPDMVSVLLDQ 288


nim     :   267 VSNVHKALDSDDIELVKLLLKEDHTNLDDACALHFAVAYCN 307
              +  + +ALD+ DIELVKL++  +  +LDDA A+H+AV +CN
Rice-3  :    33 IRRMRRALDAADIELVKLMVMGEGLDLDDALAVHYAVQHCN  155

nim     :   325 RNPRGYTVLHVAAMRKEPQLILSLLEK 351
              R P   T LH+AA    P ++  LL++
Rice-3  :   208 RRPDSKTALHLAAEMVSPDMVSVLLDQ 288


nim     :   267 VSNVHKALDSDDIELVKLLLKEDHTNLDDACALHFAVAYCN 307
              +  + +ALD+ DIELVKL++  +  +LDDA A+H+AV +CN
Rice-4  :    33 IRRMRRALDAADIELVKLMVMGEGLDLDDALAVHYAVQHCN 155

nim     :   327 PRGYTVLHVAAMRKEPQLI 345
              P G T LH+AA    P ++
Rice-4  :   215 PTGKTALHLAAEMVSPDMV 271
```

INTERNATIONAL SEARCH REPORT

Inte- nal Application No
PCI/EP 97/01218

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/82 C12N5/10 C07K14/415 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 94 16077 A (CIBA GEIGY AG ;RYALS JOHN A (US); UKNES SCOTT J (US); DELANEY TERR) 21 July 1994 see page 2, lines 20-25; page 7, lines 20-30; page 8; page 15-17; page 19, 2nd paragraph ---	1,8, 11-30
X	EMBL SEQUENCE DATA LIBRARY, 27 June 1994, HEIDELBERG, GERMANY, XP002034716 NEWMAN, T., ET AL. : "GENES GALORE: A SUMMARY OF METHODS FOR ASSESSING RESULTS FROM LARGE-SCALE PARTIAL SEQUENCING OF ANONYMOUS ARABIDOPSIS cDNA CLONES" accession No. T22612 --- -/--	1,3-5

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
9 July 1997	18.07.97

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+ 31-70) 340-3016	Holtorf, S

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/EP 97/01218

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE USA, vol. 92, July 1995, pages 6602-6606, XP002034717 DELANEY, T.P., ETAL.: "ARABIDOPSIS SIGNAL TRANSDUCTION MUTANT DEFECTIVE IN CHEMICALLY AND BIOLOGICALLY INDUCED DISEASE RESISTANCE" cited in the application see the whole document	1
X	THE PLANT CELL, vol. 6, November 1994, pages 1583-1592, XP002034718 CAO, H., ET AL. : "CHARACTERIZATION OF AN ARABIDOPSIS MUTANT THAT IS NONRESPONSIVE TO INDUCERS OF SYSTEMIC AQUIRED RESISTANCE" see the whole document	1
P,X	THE PLANT CELL, vol. 9, March 1997, pages 425-439, XP002034719 RYALS, J., ET AL . : "THE ARABIDOPSIS NIM1 PROTEIN SHOWS HOMOLOGY TO THE MAMMALIAN TRANSCRIPTION FACTOR INHIBITOR IkappaB" see the whole document	1-22,25, 28
P,X	CELL, vol. 88, January 1997, pages 57-63, XP002034720 CAO, H., ET AL . : "THE ARABIDOPSIS NPR1 GENE THAT CONTROLS SYSTEMIC ACQUIRED RESISTANCE ENCODES A NOVEL PROTEIN CONTAINING ANKYRIN REPEATS" see the whole document	1-6, 8-22,25, 28,30
P,X	EMBL SEQUENCE DATA LIBRARY, 20 January 1997, HEIDELBERG, GERMANY, XP002034869 CAO, H. ET AL.: "THE ARABIDOPSIS NPR1 GENE THAT CONTROLS SYSTEMIC ACQUIRED RESISTANCE ENCODES A NOVEL PROTEIN CONTAINING ANKYRIN REPEATS" ACCESSION No. U76707	1-6, 8-22,25, 28,30
A	WO 94 24295 A (CIBA GEIGY AG ;RYALS JOHN A (US); FRIEDRICH LESLIE B (US); WARD ER) 27 October 1994 see the whole document	26,29

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter ional Application No
PC:/EP 97/01218

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9416077 A	21-07-94	AU 6084594 A	15-08-94
		CA 2152173 A	21-07-94
		EP 0677108 A	18-10-95
		JP 8505529 T	18-06-96
WO 9424295 A	27-10-94	AU 6557494 A	08-11-94
		US 5614395 A	25-03-97